

DISCUSSION PAPER SERIES

IZA DP No. 15535

**Returns to Different Postsecondary
Investments: Institution Type, Academic
Programs, and Credentials**

Michael Lovenheim
Jonathan Smith

SEPTEMBER 2022

DISCUSSION PAPER SERIES

IZA DP No. 15535

Returns to Different Postsecondary Investments: Institution Type, Academic Programs, and Credentials

Michael Lovenheim

Cornell University, NBER, and CESifo

Jonathan Smith

Georgia State University and IZA

SEPTEMBER 2022

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Returns to Different Postsecondary Investments: Institution Type, Academic Programs, and Credentials*

Early research on the returns to higher education treated the postsecondary system as a monolith. In reality, postsecondary education in the United States and around the world is highly differentiated, with a variety of options that differ by credential (associates degree, bachelor's degree, diploma, certificate, graduate degree), the control of the institution (public, private not-for-profit, private for-profit), the quality/resources of the institution, field of study, and exposure to remedial education. In this Chapter, we review the literature on the returns to these different types of higher education investments, which has received increasing attention in recent decades. We first provide an overview of the structure of higher education in the U.S. and around the world, followed by a model that helps clarify and articulate the assumptions employed by different estimators used in the literature. We then discuss the research on the return to institution type, focusing on the return to two-year, four-year, and for-profit institutions as well as the return to college quality within and across these institution types. We also present the research on the return to different educational programs, including vocational credentials, remedial education, field of study, and graduate school. The wide variation in the returns to different postsecondary investments that we document leads to the question of how students from different backgrounds sort into these different institutions and programs. We discuss the emerging research showing that lower-SES students, especially in the U.S., are more likely to sort into colleges and programs with lower returns as well as results from recent U.S.-based interventions and policies designed to support success among students from disadvantaged backgrounds. The Chapter concludes with some broad directions for future research.

JEL Classification: I23

Keywords: returns to college, college choice, literature review

Corresponding author:

Jonathan Smith
Andrew Young School of Policy Studies
Georgia State University
P.O. Box 3992
Atlanta, GA 30302-3992
USA
E-mail: jsmith500@gsu.edu

* We are grateful to the Editors and to Andrew Barr for helpful comments and suggestions as well as to Michael Bloem for providing valuable research assistance.

1. Introduction

The early literature on human capital and the return to postsecondary investments treated college as a monolith, paying little attention to the wide differences across institutions, degrees and credentials, and educational experiences. In reality, postsecondary education in the United States and around the world is highly differentiated. There are a number of dimensions along which both higher education institutions and the type of education a student receives can vary. Students can enroll in different institution levels, which lead to different degrees, such as a Bachelor of Arts (BA) degree from a four-year college, an associate (AA) degree, diploma, or certificate from a two-year or less-than-two-year college, or a graduate degree (e.g., Master of Arts (MA), Medical Doctorate (MD) or Juris Doctorate (JD)). Even within level, postsecondary institutions vary considerably in terms of their resources, the selectivity of admissions, the composition of the student body, and the control of the institution – public, private not-for-profit, and private for-profit. Furthermore, educational experiences vary across academic majors, vocational programs, and exposure to remedial coursework while enrolled in college. The types of institutions in which one enrolls, the courses taken while enrolled, and the degree or credential earned can generate large differences in returns to postsecondary investments. Returns additionally can vary by student background characteristics.

In this Chapter, we review the literature on the returns to different types of postsecondary institutions and paths of study. The importance of identifying the returns to specific higher education investments is underscored by the increasing demand for high-skilled workers in industrialized economies such as the U.S., which typically is attributed to skill-biased technological change (e.g., Autor, Katz, and Kearney 2008; Autor 2014). The rising demand for skill raises the question of how these skills are produced and the role of different types of postsecondary institutions and programs in producing high-skilled workers. Examining the returns to different types of postsecondary investments provides insight into which types of programs and institutions are producing the skills most valued in the labor market. This information also is important to help students make optimal decisions about in which postsecondary institutions and program(s) to enroll or whether interventions are needed to increase enrollment in particular educational settings. The information on differential returns additionally can help postsecondary administrators or government agencies in deciding which programs to offer and how to allocate

funds across different institutions and programs, especially since higher education is so heavily subsidized and has the potential for social returns well beyond any private returns.¹

Rather than providing a comprehensive review of the literature, which would be too voluminous and unwieldy, we focus our review on the methodological issues that arise from the selection of students into different educational environments, the data that researchers have used to estimate the returns to specific educational investments, the economic interpretation of the parameters that researchers estimate, and the main findings. The goal is to provide readers with an overview of how researchers have analyzed the returns to different postsecondary programs and the broad conclusions (if any) reached in the literature. Our focus is on both the U.S. higher education market, which has received the most attention to date, as well as evidence from international contexts. We also highlight remaining questions that help provide direction for future research.

Our review begins with an overview of the higher education systems in the U.S. and internationally in Section 2. These systems are too numerous and the details too varied to provide a comprehensive overview of each system. Our discussion centers on the key institutional features that researchers have studied and/or used for identification. This overview helps put the discussion of the literature into context.

A core focus of the various literatures on the returns to postsecondary education is on causal identification. Students do not randomly (or exogenously) select into different higher education programs. Indeed, especially with selective colleges, there is a two-sided matching problem whereby students select colleges and colleges select students. Overcoming these selection problems is challenging and is a main focus of research on the returns to postsecondary investments. To articulate clearly the assumptions underlying the various methods researchers use and to provide a guide to the literature, we adapt the model from Altonji and Zhong (2021) in Section 3. The model shows the assumptions under which different estimators can separate the return to a given educational investment from students' pre-collegiate academic ability and their preferences over education and occupations. As the model highlights, most papers we discuss employ rather strong assumptions that are difficult to test. This does not suggest that the research

¹ This assumes that private and social returns are positively correlated with one another, which has not been demonstrated in the literature.

findings are not credible or high-quality, however it underscores the importance of future work that can relax some of these assumptions as better data and methods become available.

Our literature review begins in Section 4, where we examine the returns to enrolling in and graduating from different types of institutions. Specifically, we discuss research on the returns to a two-year college (Section 4.1) and four-year college (Section 4.2) separately, noting that sometimes the research discusses the returns between the two college levels in contrast to one another. Our discussion of the return to four-year colleges focuses on important dimensions of quality variation as well as the “match” between the academic qualifications of students and selectivity of the college. We also examine the emerging literature on for-profits (Section 4.3) and separately discuss international evidence (Section 4.4).

Section 5 discusses the research on the returns to different types of education in which students can invest. We examine vocational education, certificates, and diplomas in Section 5.1, remedial education in Section 5.2, the returns to college major in Section 5.3, and post-graduate/advanced degrees in Section 5.4.

The Chapter highlights the large differences in returns across different types of postsecondary programs and institutions. There is clear evidence that more-selective and higher-resource educational environments lead to higher returns. Given these patterns, it then becomes important to understand differences across the socioeconomic (SES) distribution in postsecondary investments. In Section 6, we discuss the U.S.-focused literature showing that lower-SES students tend to invest in less-selective and lower-resource colleges. These differences, combined with the evidence on returns discussed in Sections 4 and 5, suggest that features of the U.S. postsecondary system may actually reduce intergenerational economic mobility. We then examine the evidence on several interventions and programs that are designed to support more postsecondary investment among low-income students. We focus specifically on five policies: 1) recruiting and information simplification, 2) lowering college entrance exam costs, 3) coaching and mentoring high school students, 4) promise programs, and 5) comprehensive supports that address multiple barriers faced by lower-income students. We conclude by highlighting some potential directions for future research in Section 7.

Before we begin, we stress three points that are important to consider when reviewing the literature. First, we primarily discuss gross rather than net returns. Colleges and degree programs differ considerably in cost, and there are large cost differences across countries as well. To obtain

a comprehensive picture of different postsecondary investments, understanding both the costs and the labor market benefits are important. However, the literature focuses on gross returns, and thus we do as well. Estimating net returns and the return on investment is an important next step in these literatures.

Second, there is a difference between postsecondary enrollment and completion of a degree or credential. College completion rates are low in the U.S. (Bound, Lovenheim and Turner 2010; Bound and Turner 2011),² and different institution types and degree programs exhibit very different rates of completion. An open question in the literature is how much of the return to a postsecondary investment flows through program completion or whether there are returns to enrolling in and dropping out of educational programs. This distinction is highlighted in Section 4.

Third, our review exclusively discusses research on educational and labor market returns. These are important outcomes that have been the main focus of researchers. However, the potential returns to postsecondary investments extend beyond these outcomes in ways that are likely to be important for individuals' lifetime well-being. Outcomes such as health, family formation, civic engagement, crime, and investments in children all can be influenced by collegiate investments and have implications for private and social returns. The research on these outcomes is sparse,³ and so we limit our focus to educational attainment and labor market earnings.

2. Structure of Higher Education

In this section, we discuss the structure of the U.S. higher education system as well as some of the structures and features of higher education systems around the world. We do not provide a complete guide to higher education in every country, nor are we able to cover all the issues students, colleges, and policymakers face. Rather, we give a broad overview of the structure of higher education to provide the background for later sections where we discuss the literature on the differential returns to postsecondary institutions and programs. Understanding the returns requires knowledge of the types of institutions available to students, whom the institutions serve, and how the types of institutions interact with one another. What follows is designed to help

² Denning et al. (forthcoming) show that college completion rates have increased in recent years, however they still remain quite low.

³ See Lochner (2011) and Oreopoulos and Salvanes (2011) for reviews of research on the effect of schooling on these outcomes.

researchers and readers better understand the choices students face and the counterfactuals to students' observed choices, which is critical to estimating and interpreting the returns to an institution type or quality. It also will aid the reader on issues of generalizability and whether the estimates on returns to college investments are important in terms of magnitude or number of students impacted.

Another goal of the section is to highlight some of the institutional policies and features that researchers use in their analyses. There are too many to catalogue comprehensively, and research often exploits new policies and procedures. Instead, we focus on some promising and common institutional practices used by researchers. This allows readers to better understand how and why researchers make specific modeling decisions, the benefits and weaknesses of the approaches used, and whether there are alternative approaches that can be deployed to study the same question.

2.1. U.S. Higher Education System

Higher education in the U.S. is a decentralized market that consists of thousands of colleges, each with different attributes, policies, and governance. In any one year, there are millions of enrolled students, who have different academic achievements and varying amounts of college-related information and financial assistance at their disposal. Students also vary in their preferences over college attributes, ranging from location to fields of study to campus culture. Additionally, students differ in their knowledge of the distribution and return to these attributes. This complicated market results in an initial student-college match and ends in a dropout decision, transfer to another institution, or a postsecondary credential.

Much has been written about the structure of higher education in the U.S. in papers that vary by discipline and scope (Hoxby, 1997; Winston, 1999; Goldin and Katz, 1999; Lowry, 2001; Knott and Payne, 2004; McLendon et al, 2009; Tandberg, 2010). There also is a substantial body of work in economics on how the structure of higher education has evolved over time and the implications for students and the labor market (e.g., Goldin and Katz, 1999; Hoxby, 2009; Lovenheim and Turner, 2017). In this subsection, we provide a few basic facts on the types of institutions in the U.S. We also highlight how the types of institutions differ from one another, including the students who enroll in them, institutional financial resources (tuition, aid, and expenditures), and degrees awarded.

2.1.1. Basic Structure and Students Served

There are many ways to categorize postsecondary institutions in the U.S., but there are a few primary distinctions relevant for the vast majority of students and on which researchers focus. First, there are degree-granting and non-degree-granting colleges, the former of which awards associate (AA) or bachelor's (BA) degrees or both. Those same institutions also may offer other credentials, such as diplomas and certificates or post-baccalaureate degrees. As seen in Table 1, in the 2019-20 academic year, there were 3,757 degree-granting institutions in the U.S.⁴

Second, there are different *levels* of institutions, which are defined by the U.S. Department of Education as four or more years, at least two-years but less than four-years, and less-than-two-years. Of the 3,757 degree-granting institutions in the U.S., 2,476 (66 percent) are four-year institutions that offer a BA degrees or higher and 1,269 (34 percent) are two-year institutions offering AA degrees and other certificates. Similarly, total enrollment in four-year colleges is roughly two-thirds of total enrollment, regardless of the measure of enrollment used. Four-year colleges tend to award many more BA degrees than AA degrees, whereas two-year colleges almost exclusively award AA degrees. The lines between these levels are increasingly blurred, whereby there are increasingly more colleges that offer both AA and BA degrees.

Other than degrees awarded, these levels of institutions differ by both the students they serve and the resources at their disposal. The students served at two-year colleges are more likely to be from economically disadvantaged and underrepresented populations. For example, 46.8 percent of four-year college enrollees receive a Pell Grant, a measure of financial need, compared to 59.4 percent at two-year colleges. And 33.8 percent and 42.2 percent of enrollees at four-year and two-year colleges, respectively, are underrepresented minorities (URMs).⁵ Tuitions and expenditures per students at four-year colleges also are higher than at two-years, the former of which contributes to differences in who enrolls and the latter of which contributes to differences in graduation rates (Webber and Ehrenberg, 2010). For example, the average published tuition and fees for four-year colleges is \$20,977, compared to \$6,486 for two-year colleges. Instructional expenditures per student are nearly double at four-year colleges.

⁴ There are over 1,700 institutions that are not degree-granting (authors' calculations from the Integrated Postsecondary Education Data System (IPEDS)). They are mostly for-profits and only enroll 1.8 percent of students as of fall 2018 (de Brey et al., 2021, Table 303.20). These measures from IPEDS only consider colleges that participate in federal aid programs, which is the vast majority of colleges, especially non-profit colleges.

⁵ Underrepresented minorities are typically defined as African Americans, Hispanic Americans, and Indigenous Peoples. First generation college students also sometimes are included in this definition.

Third, institutions differ in their *control*, which is, generally speaking, whether an institution is publicly or privately operated and governed. Public colleges are governed by state and local governments through a governance board.⁶ Because these colleges rely so heavily on state appropriations and are under government control, they are limited in the tuition they can charge, their admissions policies, and in the general operation of the university. There are 1,636 public degree-granting institutions (44 percent) and 2,109 private institutions in the U.S.⁷ Although not shown, the percent varies by state. For example, in Indiana, only 22 percent of colleges are public, while Maryland and Wyoming are 56 and 89 percent public, respectively.⁸ Occasionally, we refer to public two-year colleges as community colleges. More frequently and with a slight abuse of precision, we refer to public two-year colleges as two-year colleges. There are relatively few private not-for-profit two-year colleges, and we make certain to differentiate with for-profit two-year colleges when appropriate.

Fourth, within private institutions there is a distinction between not-for-profits and for-profits. Of the 2,109 private institutions documented by the federal government, 1,461 (69 percent) institutions are not-for-profit. However, many for-profits do not participate in the federal aid programs and therefore are not subject to reporting. Cellini and Goldin (2014) estimates that about half the for-profits are not captured by federal data but also note that most of these institutions are small and do not grant degrees.

Despite the smaller number of public institutions, they serve approximately three-quarters of enrollees. Table 1 shows that along a limited number of student characteristics, such as percent who receive Pell Grants and percent URM, public and private colleges are quite similar on average. However, private for-profit institutions have much higher rates of enrollment by Pell Grant recipients and URMs than either public institutions or private not-for-profits. The disparities in tuition between these different types of institutions is the focus of a substantial amount of research. For in-state public colleges, the average published tuition and fees is \$6,311, compared to \$28,257 for private not-for-profit and \$15,832 for private for-profit colleges. These correspond to instructional expenditures per student of \$8,567, \$11,213, and \$5,151, respectively, for those different control categories. We also note that there are additional and meaningful differences in

⁶ The U.S. Federal government also operates a small number of military academies.

⁷ More information on public colleges, their governance, history, and policy issues can be found in Lovenheim and Turner (2017) or Bound et al., (2019).

⁸ Source: de Brey et al. (2021), Table 317.20.

enrollees, tuitions, and expenditures within each level and across controls. Uniquely, the difference in average published tuition and fees between two-year and four-year for-profit colleges is quite similar, compared to the nearly doubling of tuition across levels among public and not-for-profit colleges.

The above statistics show some of the raw differences across institution types, the number of students served, and students' backgrounds. Similar patterns persist in college choice analyses that use regression models to separate attributes of colleges and students that are correlated with one another. Specifically, economically disadvantaged students are much more likely to enroll in two-year colleges than four-year colleges, public colleges over private colleges, and for-profits over not-for-profits than their relatively economically advantaged counterparts, even after accounting for prior academic achievement, other demographic characteristics, and geography (e.g., Long, 2004; Reynolds, 2012; Cellini and Darolia, 2017). The same is often true for underrepresented minorities (URM), especially in relation to for-profits (Cellini and Darolia, 2017). However, we note that race and ethnicity are often related to prior academic achievement and parental income in complicated ways that can translate into different patterns when one controls for these factors (e.g., Reynolds, 2012).

2.1.2. Application and Admission

Because the U.S. system of higher education system is decentralized, enrollment into any of the thousands of colleges is almost always handled independently by each institution. Approximately half of the colleges have no admission criteria and are considered "open enrollment," but this varies by institution type. For example, 25 percent of four-year colleges have no admission criteria compared to 92 percent of two-year colleges. And 65.3 and 36.8 percent of public and private colleges, respectively, have no admission criteria.⁹ These differences in criteria partially determine the students they serve, which we discuss below.

Among the four-year colleges that do have selective admissions, there are many steps and decisions students must take in order to enroll (e.g., Klasik, 2012). We highlight a few defining characteristics of this system. First, students have to send an application to the college. The details of the application are determined by the college (or sometimes the governing body) and can be as simple as a high school transcript. More often than not, the application is intended to allow the college to gauge the students' academic and non-academic fit within the institution.

⁹ Source: de Brey et al. (2021), Table 305.40.

Second, many colleges that are not open enrollment require students to take the ACT or SAT, which are the two college entrance exams in the U.S. Unlike other parts of the world, the exams are offered multiple times per year and can be retaken.¹⁰ In some states, the ACT and SAT are mandatory for all public high school students (see Section 6). The exams are a measure of academic preparedness on mathematics, reading, and writing.

Third, application requirements can vary dramatically across institutions. Some institutions base admission purely on academics, such as high school courses taken, GPA, and ACT or SAT scores. Other colleges use “holistic” admissions that require applicants to detail some or all of their extracurricular activities, academic and non-academic achievements, letters of recommendation, interviews, essays, and field of interest. The fact that some colleges require different pieces of information, along with separate application fees, has resulted in a few colleges entering application consortia (e.g., the Common Application). The application process for a highly-selective college is a very complicated, expensive, and unwieldy process that some researchers show can influence application and enrollment decisions, particularly of students who are from less-advantaged backgrounds (e.g., Pallais, 2015; Hoxby and Turner, 2013).

While many students simply apply to the one college they are most interested in,¹¹ often with either open enrollment or relatively open access, some students’ applications are less certain. As such, conditional on applying to one college, the typical student applies to almost three colleges. However, there is a long right tail for those applying to the most selective colleges with low admit rates.¹²

Once students send in their applications, they are evaluated separately by each institution and students can be offered admission to any or all of the colleges to which they apply. Often times, students will receive a financial aid offer along with a letter of admission that states how much of the tuition the student (and her family) is expected to pay. In turn, students decide which college they choose to enroll in among the ones extending admission. As highlighted in Dynarski and Scott-Clayton (2006), the financial aid application process itself is quite complicated and opaque, which creates access barriers for the most disadvantaged students. That students do not

¹⁰ Retaking the ACT and SAT is done at unequal rates and also contributes to inequality in racial and economic four-year college enrollment rates (Goodman, Gurantz, and Smith, 2020; Bloem, Pan, and Smith, 2021).

¹¹ There are also opportunities for students to apply and be accepted early, before other applications are due. This alternative strategy and its implications is the subject of some economics research (e.g., Avery, Fairbanks, and Zeckhauser, 2009; Avery and Levin, 2010) but is worthy of more.

¹² Authors’ calculations using High School Longitudinal Study of 2009.

know their financial aid packages prior to applying makes it difficult for students to choose selective institutions based on price.

2.1.3. College Selectivity and Inequality

There is a substantial amount of research on the returns to college quality, with “high-quality” institutions often referred to as selective or elite colleges. Research on the returns to college quality tends to focus on four-year colleges and hence the within-level relative quality effects, but some research also has examined quality effects more broadly among the entire higher education market (e.g., Dillon and Smith, 2020). However, there is no consensus on the definition of quality and moreover, different researchers use the same variables to be inputs, outputs, control variables, and measures of quality. These variables and sometimes measures of quality include the demographic and economic composition of the student body, average academic ability of enrolled students, expenditures per student, and completion rates. While many of these variables are sensible measures of quality, the clear relationship between college type, student characteristics, and positive outcomes makes estimating the returns to college quality a difficult task.

Table 2 shows the relationship between colleges’ Carnegie Classification and several college characteristics. The Carnegie Classification groups colleges into 15 types, based on whether they are two-year or four-year, the fraction of part-time versus full-time enrollees, the fraction of transfers-in (four-years only), and a measure of selectivity (four-years only). We aggregate across groups so as to compare their three measures of four-year selectivity, which are based on the ACT and SAT scores of first-year enrollees:

- More selective - 80th to 100th percentile of selectivity among all baccalaureate institutions;
- Selective - 40th to 80th percentile of selectivity among all baccalaureate institutions;
- Inclusive - these institutions either did not report test score data or the scores indicate that they extend educational opportunity to a wide range of students with respect to academic preparation and achievement.

Although researchers use a range of measures for selectivity or quality, the patterns follow that of Table 2 – the more selective institutions compared to the more inclusive institutions serve fewer economically disadvantaged students and URMs, the admissions process is more competitive, and

they have higher expenditures per student, graduation rates, and posted tuition and fees (although institutional financial aid also is more generous).

We highlight a few specific statistics in Table 2 to demonstrate the aforementioned disparities. First, there are relatively more public colleges than private not-for-profit colleges outside the more selective four-year colleges. Additionally, although most of the for-profit colleges are two-year institutions, almost all of the four-year institutions that are for-profit are inclusive. Second, the average “more selective” institution has 25.8 percent of students receiving Pell Grants, compared to 40.7 and 57.7 percent for selective and inclusive institutions, respectively. The inclusive four-year institutions are quite similar to the two-year institutions in regard to Pell Grant receipt as well as percent URM. This demonstrates a feature of the system that is highlighted in much of the literature—many students’ decisions are whether to enroll in a public two-year or relatively inclusive four-year college.

Third, more selective institutions are the ones with very competitive admissions processes that receive lots of applications and have low admit rates. This corresponds to average SAT scores that are about 180 and 270 points higher than selective and inclusive four-year institutions, respectively. Peers are potentially part of the treatment effect of attending a college, and so the academic preparedness of matriculants may be important.

Fourth, the average published tuition at the more selective institutions is double that of the inclusive institutions and four-times that of a two-year college.¹³ This pattern is evident even at public colleges, although to a lesser extent. These relatively higher sticker prices correspond with much higher expenditures per student, perhaps contributing to the steep relationship between selectivity and graduation rates, dropping from 75.4 to 56.7 to 40.3 percent across selectivity groups.

Table 2 provides an overview of the variation across institutions in their resources, student body composition, and educational outcomes. This heterogeneity has been a core focus of the returns to postsecondary education literature. Given the complex and decentralized application process discussed above, a central challenge is on disentangling unobserved determinants of selection from causal estimates of college type on educational and labor market outcomes. Section 4 provides a detailed discussion of how researchers have addressed these selection problems to

¹³ These published tuition prices are not necessarily reflective of the final price students pay, net of financial aid and institutional grants (Ma, Pender, and Libassi, 2020).

identify the effect of enrolling in different types of institutions, while Section 5 focuses on the literature that estimates the return to different types of programs.

2.2. Higher Education Systems Around the World

There are nearly as many distinct higher education systems around the world as there are countries, each one different from another in some or many ways and all different from the U.S. system. In this subsection, we highlight some of the features of higher education systems around the world that appear in the economics literature. The goal is not to describe every feature of every country. Rather, we focus on a few countries that appear prominently in the literature and highlight some features of these systems to which researchers rely. A summary of some of the central features of the higher education systems around the world are included in Table 3.

2.2.1. Institution Types

Across the world, there are many types of institutes of higher education, most of which have an analogous college type in the U.S. that we described above. At the broadest level, perhaps the biggest difference from the U.S. is the distribution of these types of institution within a country and their degree of autonomy. We detail some important facts about several countries' institution types in Table 3.

Public colleges and universities tend to be the dominant type of institution in most countries, in terms of quantity and often prestige. For example, most colleges are publicly or state-funded in Australia, the United Kingdom (U.K.), and Norway, while the most prestigious colleges that have rigorous admissions standards in Colombia and India are public. But almost all countries that have received attention in the literature also have private institutions, sometimes offering more of a vocational education or something more akin to a two-year college in the U.S.

Additionally, public colleges around the world differ in the administrative levels at which they are governed. Some colleges are financed and operated by the national government, which is largely the case in Australia and the U.K. On the other hand, some countries' public universities, such as India's, are more similar to the U.S. and funded by state or provincial governments, not the federal government. And finally, there are hybrids of national- and state-funded colleges (e.g., India).

Most countries also segment their postsecondary systems by the types of degree or occupation associated with the intended degree. The largest such example is countries' delineation of vocational colleges. Even within vocational colleges there can be finer gradations. For

example, in Colombia, there are public and private technical colleges and sub-baccalaureate technical colleges. Frequently, these different types of institutes have different governing bodies, such as national agencies that focus on one type of an education or institution over another. In Europe, there are polytechnic colleges that focus on vocational training, often with four-year programs.

Finally, there are also for-profit colleges around the world, but similar to the U.S. they are relatively few. Most countries do not have these types of schools, with some notable exceptions. In Brazil, a country with relatively low college enrollment rates, nearly one-third of enrollees are at a for-profit institution (Salto, 2014).

While this summary of institution type does not dramatically depart from the U.S. framework, it does offer insights into some of the ways in which the returns to different college types has progressed in other countries and how it is framed. For example, in some contexts, it is impossible to differentiate the returns to public institutions from quality or prestige. Additionally, there are often more distinct tracts, agencies, and governing bodies within certain countries that make the opportunities for students to consider multiple types of institutions difficult. Unlike the U.S., this may simplify the choice set that students consider.

2.2.2. College or College-Major

The process by which major is chosen and the timing may have important implications in the labor market, and these processes vary across countries. First, field of study has been shown to be at least as important as college quality in terms of labor market returns in a variety of settings around the world, including in the countries that require major to be chosen at the time of application (e.g., Kirkeboen, Leuven, and Mogstad, 2016). Second, choosing a major prior to starting college likely means students are more likely to stick with those majors than in the counterfactual world with fewer barriers to switching or more time to declare a major. The cost of switching majors can be very costly. It plays a role in initial major choice and the probability of switching, as modeled in Altonji, Arcidiacono, and Maurel (2016). Research shows that choice of major can be influenced by the mere exposure to different coursework (e.g., Fricke, Grogger, and Steinmayr, 2015), and switching majors is quite common in settings that allow students to do so more freely and as new or updated information becomes available (e.g., Arcidiacono, 2004; Stinebrickner and Stinebrickner, 2014).

One of the most common differences between the U.S. and other countries' higher education systems is when students declare a major. In the U.S., students often apply to a four-year college, perhaps stating their intended major. However, it usually is easy to switch majors in the first two years of enrollment. In other countries, students frequently apply to a college and a major simultaneously. This is true in Brazil, Chile, Colombia, Croatia, France, Italy, Norway, and Sweden, to name a few. We detail application and admission processes below, but in broad strokes, students' decisions are based on preferences over colleges, majors, and the combination of the two. Whether they are admitted depends on the college-major combination. Of course, many colleges have procedures by which students can switch majors, but they often come with costs and uncertainty. As a result, major switching is much less common in countries where students apply to an institution and major simultaneously.

Some countries are more similar to the U.S., whereby students apply directly to a college and students sort into majors after matriculation – Australia, Canada, and the U.K. are examples. However, it is common for colleges to specialize in particular areas of study or majors, and so applications to the college are implicitly choosing majors. For example, there are sets of engineering schools in India and vocational schools in almost every country.

2.2.3. Tuition

Tuition and financial aid are arguably some of the most frequently researched topics in the economics of higher education. One reason is because tuition and financial aid is a clear policy lever at the disposal of colleges and policymakers to achieve desirable educational outcomes. Related, higher education is often provided or heavily subsidized by local and national governments, which makes tuition setting and aid an important function of government agencies. The empirical research, especially in the U.S., is clear – financial aid (and lower tuition) increases college access and enrollment, especially for lower-income students (Deming and Dynarski, 2009; Dynarski and Scott-Clayton; 2013; Page and Scott-Clayton, 2016).

Because students may face credit constraints, along with disparities in who attends college across the globe, tuition and aid policies within and across countries have the potential to be of great importance to individuals and governments. First, these policies are likely to impact the extensive margin of whether to attend college. Second, they may impact the intensive margin of the type and quality of the college in which a student enrolls. Third and the focus of this Chapter,

these decisions on whether and where to enroll have an impact on the return to college in the labor market.

The most common feature of tuition across several countries that have substantial numbers of public and private colleges is that tuition at public colleges is lower than at private colleges. For example, in Colombia, tuition at public colleges is roughly half to one-third that of tuition at private colleges (González-Velosa, 2015). In some countries, public colleges are entirely free, such as in Brazil and Norway. College used to be free in the U.K., but in recent years students have faced increasingly large fees (Murphy, Scott-Clayton, and Wyness, 2019). In other cases, like Chile, public colleges (and some privates) are free only to lower-income students. These variations in tuition policies over time and across subgroups present opportunity for additional research on the impacts of tuition policies on college enrollment and on the returns to enrollment (e.g., Murphy, Scott-Clayton, and Wyness, 2019).

Another common feature of tuition policies around the world is that tuition is often college-major specific. For example, in Australia and Chile, tuition depends not just on the college but also on the major within the college. In some countries, since tuition, applications, and admissions are college-major specific, tuition has the potential to influence field of study and occupation as well as produce inequitable access to certain fields due to financial constraints. Additionally, some countries charge a uniform tuition across majors but have additional fees for different majors,¹⁴ often dwarfed by tuition.

Finally, tuition in some countries is quite expensive, regardless of whether the colleges are public or private. South Korea is an example, especially compared to some countries like France and Italy. The U.K. and Australia also have high tuition (or fees), however, they are unique in that they do not require students to pay their tuition until they leave college. Rather, students accumulate debt and in the case of Australia, repayment is tied to income so students with low or no income do not have to make payments and, unlike some students in other countries, avoid defaulting on their debt.¹⁵

2.2.4. Application Process and Centralization

¹⁴ Major-specific fees also appear in the U.S. But they are distinct from published tuition, which is typically the same for all majors within a college. Stange (2015) and Andrews and Stange (2019) offer some examples.

¹⁵ Income-driven repayment is an option in the U.S. but it is rarely chosen by students, despite the potential benefits, and one needs to opt into it (Cox, Kreisman, and Dynarski, 2020).

The college application process has important implications for a country's higher education system. First, the process is a sorting mechanism, which almost uniformly across contexts makes certain groups of students more likely to apply and enroll in certain types of colleges (e.g., Dillon and Smith, 2013). Application processes can be complicated, which can perpetuate inequality because of informational asymmetries, financial constraints, peer effects, or inertia. Second, the process is a matching mechanism that matches students with different academic backgrounds, non-academic attributes, and preferences to colleges (and often majors). We later discuss match quality in the U.S. setting, but both theoretical and empirical findings suggest that the student-college academic match can contribute to student college completion and labor market earnings (e.g., Chade, Lewis, and Smith, 2014; Fu, 2014; Dillon and Smith, 2020). But match quality extends beyond academics, sometimes referred to as "fit," and may impact student outcomes as well. Overall, the application process is the process that dictates who goes where and who ultimately has access to the types of institutions that have the low or high returns that we document in this Chapter.

At the extremes, there are entirely centralized and decentralized higher education markets, and the application processes follow suit. Decentralized systems, like the U.S., require students to apply to each college separately, often paying an application fee at each college, receiving independent admissions decisions, and students choosing to enroll in one college.¹⁶ These entirely decentralized application systems are rare in the industrialized world. Rather, there is usually some coordination, either among state agencies or colleges. For example, many countries have a coordinated application system where students enter their relevant information that is considered by all colleges designated by the student. The system may be for a subset of colleges, such as the "traditional" and older colleges in Chile (CRUCH), colleges in a specific geographic region, such as state-specific applications in Australia, or for all colleges in the country, as in Norway.

The most centralized systems employ a matching algorithm, such as the sequential dictatorship method, where students rank-order a set of colleges or college-majors and in turn, colleges or college-major rank applicants, often on test scores and high school grades. Some countries only allow for up to five colleges (e.g., U.K.), while others allow for up to 15 (e.g.,

¹⁶ There are some efforts to centralize the system in the U.S., like the Common Application or some state-specific application procedures to the public colleges, but they still require separate application fees and often supplemental material.

Norway).¹⁷ Students then receive a single offer from their highest-ranked college that is willing to admit the student. This type of matching algorithm and its variants are used in Chile, Norway, Colombia, Croatia and to some extent, China. These highly centralized systems only give students a single enrollment option among the participating colleges. In contrast, the U.K. has a coordinated and centralized application system but colleges make their admission decisions independently, such that students can receive multiple offers of admission.

Students are typically forced to rank their colleges, usually in a strategy-proof order (Svensson, 1999) in most centralized systems, which can prove especially valuable to researchers. This means that researchers see students' preferences over colleges, which is not possible in decentralized systems like the U.S. or Italy where researchers may see an unranked set of applications. The ranking of colleges provides researchers great detail about how students make decisions but perhaps more importantly, provides a clear counterfactual – the next-ranked college to which the student would have been admitted in the absence of the observed admission. Since these centralized systems often have simple admission rules so colleges also can rank students, such as test score cutoffs, the counterfactual is directly observed and not inferred.

Of course, each country has its own unique quirks, many of which have the potential to sort and match students with colleges in impactful ways. We list a few of them in the application process column of Table 3. For example, in the U.K., once conditionally admitted, students choose a top college and an “insurance” college, both of which require a sufficient level of performance to be officially admitted. And in South Korea, early admission is extremely popular.

Over and above these application features, especially those in more centralized systems, there are usually some colleges that operate independently or outside the centralized process. These are often relatively open access institutions and tend not to be over-subscribed. As in the U.S., assessing the returns to these institutions that do not have strict application rules is challenging for researchers, who cannot disentangle the effect of student attributes from the causal effect of enrolling in the college.

2.2.5. Admission Criteria

Admission criteria vary widely across the world and are intertwined with the application process. This is clear from the above description of the centralized systems, to which many countries subscribe. Admission criteria are potentially valuable tools for researchers since they

¹⁷ China allows for up to five colleges and 3-5 majors within each college.

can help assess the returns to college type and quality by imposing rules and decisions that sort very similar students into one college over another. This remains a challenge for open access institutions.

One common theme across countries is that admission is heavily weighted towards college entrance exam scores, sometimes exclusively and sometimes in conjunction with high school grades. In several countries, high school students all take the same college readiness exam and that exam is the primary or sole factor in admission. The unidimensional factor allows for the centralized systems to rank students on exam scores with little work on the end of each college. However, similar exams also are used in more decentralized systems, like India and South Korea.

One appealing byproduct of these heavily-weighted national exams is that most students in a country take the exam. Sometimes the exams are required to graduate from high school (e.g., France). This gives researchers and policymakers a metric with which to assess the entire distribution of scores across the country, even among non-college goers. Other countries, like the U.S., suffer from selective college entrance exam taking, which creates a barrier to college enrollment (e.g., Goodman, 2016) and also makes it difficult to assess how the full distribution of students sorts into college, college type and quality, or no college at all.

Along with nationally-administered exams, two related features appear prominently in admissions systems around the world. First, high school grades are often used in admission alongside exam scores. Even though this introduces a second factor, it is numeric and still allows for colleges to rank students on academic achievement, including in a centralized system. Second, in more decentralized markets, many colleges have their own college-specific entrance exams or major-specific exams that can be used across colleges. These specialized exams are useful in that they are frequently employed as a central factor in admission and consequently, can be used in assessing the returns to college (e.g., Anelli, 2020). However, they also introduce costs to students and likely impact who takes the exams and who has access to these types of colleges.

Some countries make use of quotas or affirmative action type policies in admission as well. From a researcher's perspective, quotas and similar policies help draw some students in and keep some students out of particular colleges (e.g., Black, Denning, and Rothstein, Forthcoming). Additionally, they are implemented in a variety of ways, which creates different and plausibly exogenous variation in college enrollment at different margins and contexts. There are regional and national quotas. For example, China has provincial quotas that dictate exactly how many

students from a particular province can attend a certain college and/or major. South Korea has a complicated and changing set of regional and national quotas that are college-specific. As an alternative, Brazil has race-based quotas and numerous other places, even in the centralized processes, have preferential treatment for underserved populations (e.g., people from the north in Norway, children of military casualties in China).

Finally, despite focusing on these large and clear admission policies, there are many contexts within and across countries where admission is based on additional factors. While some colleges heavily weight academic metrics, many colleges also require personal interviews, engagement with extracurricular activities, or letters of recommendation – something akin to “holistic” review in the U.S. These criteria often are vague and varied, making it difficult for researchers to observe and exploit the true admission policy.

3. Estimating the Returns to Postsecondary Degrees

One of the core challenges with estimating the return to postsecondary degrees is causal identification (Card, 2001). Students select into different types of colleges and different degree programs based on a host of factors that are challenging to measure, such as preferences for different types of occupations and educational environments, taste for different subjects, pre-collegiate academic achievement (often referred to as “ability”), and the disutility of effort. One can proxy for these taste and ability parameters, but even a rich dataset is unlikely to adequately control for the selection of students into different educational institutions and programs.

Researchers address this selection problem in many ways. A main focus of this Chapter is on understanding the approaches researchers take to overcome selection and the assumptions under which these approaches identify causal impacts of degree programs on educational and labor market outcomes. We adapt the model from Altonji and Zhong (2021) as a guide to clarify the assumptions embedded in different approaches to estimating the return to postsecondary education. The original specification of this model focuses on the returns to graduate school education, but it is a more general framework for articulating the assumptions needed to justify different approaches used in the literature. We will return to the original specification of this model in Section 5.4 when we discuss the return to graduate degrees.

For the moment, we restrict our attention to estimating the returns to a college degree, however it is straightforward to extend this model to examine postsecondary enrollment rather

than completion. Let i denote an individual and t denote the calendar year. For simplicity, t also indexes the year since obtaining a degree. Postsecondary degrees are indexed by c , where $c=0, 1, \dots, C$ are the different degrees an individual can earn. For example, an AA in nursing, a BA in mathematics, and a certificate in Information Technology would each be a separate value of c . The variable c also can denote degrees from different types of institutions, such as a BA in economics from a public versus a private four-year university, or an AA in English from a public two-year college versus a for-profit college. Individuals who do not have any post-secondary training are coded as $c=0$, and an individual's occupation is given by $j=1, \dots, J$.

There are two sources of individual heterogeneity in this model: ability (A) and preferences (Q). The vector A_{it} is a measure of individual ability. It refers to all characteristics of an individual that are correlated with the earnings of workers in occupation j given education c . An individual's wages are then given by $w_{ijct}=w_{jct}(A_{it})$.¹⁸ The vector Q_{it} refers to preferences for both the type of education one receives (c) and the choice of occupation given one's education level (j). For example, a worker could have a strong interest in computers that would induce her to enroll in an IT vocational program and then work in an IT-related job. Preferences and ability are related, as one may be more productive in an occupation in which one has more of an interest. Conceptually, this model distinguishes between these two factors, with any productivity effect of preferences embedded in A (and thus in wages) so that Q only affects decisions over degree programs and occupations.

Although it is not stated explicitly in most papers, the parameter of interest when researchers study the returns to postsecondary degrees is the treatment effect on the treated (TT). This is the effect of the educational program on earnings among those who enroll in the program (or earn a given degree). Let $\bar{w}_{0t}|C_{ct}$ be the mean earnings of what people would have earned had they not obtained a given postsecondary credential. Hence, $\bar{w}_{ct}|C_{ct}$ is the mean earnings in year t of those who obtain degree c , and $p_{ct}(j|A_t, Q_t)$ is the probability of choosing occupation j in time t given A_t , Q_t , and c . Finally, let $dF_t(A_t, Q_t|C_{ct})$ be the conditional density of A and Q , which is conditional on the educational choices individuals make. We now can express mean wages as follows:

¹⁸ Note that this function does not restrict how A affects wages, so it can interact with both occupation and educational attainment.

$$\bar{w}_{ct}|C_{ct} = \sum_j \int_{A,Q} p_{ct}(j|A_t, Q_t) w_{cjt}(A_t) dF_t(A_t, Q_t|C_{ct}) \quad (1)$$

$$\bar{w}_{0t}|C_{ct} = \sum_j \int_{A,Q} p_{0t}(j|A_t, Q_t) w_{0jt}(A_t) dF_t(A_t, Q_t|C_{ct}), \quad (2)$$

where equation (2) is the counterfactual earnings of those who choose c had they not obtained a degree. The treatment on the treated effect is:

$$TT_{ct} = \bar{w}_{ct}|C_{ct} - \bar{w}_{0t}|C_{ct} \quad (3)$$

The first part of this expression is observed in the data: it is the mean earnings of individuals who obtain degree c . The second part is the unobserved counterfactual earnings that needs to be estimated. For ease of exposition, we have written the model such that the counterfactual is for those who obtain no degree. One could also adapt this model to compare different degree holders, for example comparing the return to earning a BA versus an AA degree or comparing the return to earning a BA in economics versus English.

This framework highlights that the returns to a given degree come from two sources. The first is how obtaining degree c affects the distribution of occupations selected, conditional on A and Q . The second is changes to the potential earnings of workers within a given occupation. Hence, obtaining a particular degree could alter earnings by changing the types of occupations for which a worker is qualified. Obtaining a nursing degree allows one to be a practicing nurse, and part of the return to this degree is driven by earnings of nurses relative to what workers otherwise would do. Postsecondary degrees also may make workers more productive conditional on their occupation by raising their level of human capital. For the most part, studies of the returns to postsecondary degrees do not distinguish between these two mechanisms. However, they are conceptually distinct, and the relative importance of each is likely to vary substantially across degrees.

This framework allows us to articulate clearly the identifying assumptions required for common estimators used in the literature. For simplicity, consider a three-period model, where period 1 is prior to college, period 2 is during the period of college enrollment (for those who enroll), and period 3 is post-college. We assume people work in period 2 if not enrolled in college.

First, consider a cross-sectional estimator, which compares the earnings of those with a given degree to those who do not enroll. The treatment effect on the treated is:

$$TT_{ct}^{cs} = \bar{w}_{c3}|C_{c3} - \bar{w}_{03}|C_{c3}, \quad (4)$$

where $\bar{w}_{03}|C_{c3}$ is counterfactual earnings that is estimated using the earnings in period 3 of non-enrollees: $\bar{w}_{03}|C_{03}$. Plugging this expression into equation (2) shows that $\bar{w}_{03}|C_{03}$ is an accurate counterfactual for $\bar{w}_{0t}|C_{ct}$ when $F_3(A_3, Q_3|C_{03}) = F_3(A_3, Q_3|C_{c3})$. In words, the distribution of ability and preferences need to be the same across those who invest in a given degree program and non-attendees. This is a strong assumption, as students sort into degree programs as a function of both ability and preferences. Even with a rich dataset that has extensive controls for A and Q , it is unlikely that one can completely account for the myriad ways in which students who obtain a given degree differ from those who do not enroll in college.

One also could compare the return to a degree relative to another degree. This is common in the returns to college major literature, where researchers typically compare the return to a given major (e.g., economics) relative to another major (e.g., liberal arts). Letting c^1 denote the first degree and c^2 denote the comparison degree, the treatment effect on the treated can be written as: $TT_{ct}^{cs} = \bar{w}_{c^13}|C_{c^13} - \bar{w}_{c^23}|C_{c^13}$. Plugging into equations (1) and (2) shows that this approach identifies the return to degree 1 relative to degree 2 only when the distribution of ability and preferences among those who obtain degree 2 is the same as among those who obtain degree 1 (i.e., $\bar{w}_{c^23}|C_{c^23} = \bar{w}_{c^23}|C_{c^13}$). This may or may not be a reasonable assumption, depending on the similarity of those who select into these different degree programs and the set of controls in the data. For example, estimating the return to a physics versus a chemistry degree among Georgia State University students requires weaker assumptions because the students already have been selected by Georgia State and have chosen STEM majors that require similar technical skills and that can lead to similar occupations. In contrast, comparing the return to a BA in chemistry at Georgia State versus an AA in chemistry at Tompkins County Community College would require much stronger assumptions.

Because of the strong assumptions embedded in cross-sectional estimators, researchers seek other ways to account for selection into degree programs as a function of ability and preferences. One prominent method is to find an instrumental variable that affects program

enrollment in a manner that is orthogonal to ability and preferences. Let Z_{it} be such an instrument, then the treatment effect on the treated is:

$$TT_{ct}^{IV} = \bar{w}_{c3}|C_{c3}, Z_{c3} - \bar{w}_{03}|C_{c3}, Z_{03} \quad (5)$$

This is identified under the assumption that $\bar{w}_{03}|C_{03}, Z_{03} = \bar{w}_{03}|C_{c3}, Z_{03}$. That is, conditional on Z , the distribution of ability and preferences are the same on average and thus the earnings among the non-enrollees is an accurate counterfactual for the earnings of those who enroll. This is a way to formally state the exclusion restriction: the instrument must affect earnings only through its effect on degree enrollment/completion.

One of the most prominent instruments in the literature is distance to a four-year college (Card, 1995). The exclusion restriction needed to identify this model is that distance to a four-year college *only* affects earnings by making it more likely that those who live closer enroll in college. Conditional on the controls in the model, if distance to a college is correlated with variation in ability or preferences of college-age students, then the exclusion restriction fails. Those who live closer to a college may be systematically different from those who live farther, in part because local economies are impacted by the existence of postsecondary institutions. We discuss this particular instrument and its variants in more depth in Section 4.

Regression discontinuity (RD) is another type of method used with increasing frequency by researchers. In order to use this method, one needs a discontinuous assignment rule to different degree programs. These are relatively rare in practice and the discontinuities are almost always used as instrument for the endogenous enrollment variable. Examples include high school GPA or SAT cutoff rules for admission to a four-year college (Hoekstra, 2009; Zimmerman, 2014; Goodman, Hurwitz and Smith, 2017), high school GPA or family income cutoffs for access to financial aid (Bettinger et al., 2019; Scott-Clayton, 2011; Scott-Clayton and Zafar, 2019), and GPA/admission score cutoffs for access to specific college majors (Hastings, Nielson, and Zimmerman, 2013; Kirkeboen, Leuven, and Mogstad, 2016; Andrews, Imberman and Lovenheim, 2017; Bleemer and Mehta, forthcoming). Under the assumption that the distribution of student characteristics moves smoothly through the cutoff, the running variable acts as a control function that fully accounts for differences in the distribution of ability and preferences across the treatment threshold.

Many researchers also exploit policy variation in a difference-in-differences setting, and these policies often can be interpreted as instruments. Some examples include free community college tuition, implementation of “percent plans” for college admissions, and college entry/exit in a local area. In these cases, the exclusion restriction consists of the traditional assumptions underlying a difference-in-difference model: untreated units must be an accurate counterfactual for treated units.¹⁹

Importantly, IV approaches to estimating the return to different types of college degrees identify a local average treatment effects (LATE). If the complier population differs substantively from the overall population of interest, the LATE can differ from the treatment effect on the treated.²⁰ This can make IV studies difficult to compare to other studies. For example, Hoekstra (2009) uses SAT admission cutoffs for a flagship university to identify the effect of attending the state flagship on earnings. This identifies the effect of flagship university enrollment among students who are the least academically qualified. More academically advanced students may experience different returns, which would make the treatment effect on the treated differ from the LATE in this context.

A final approach used in the literature is individual fixed effects with longitudinal data. The benefit of this approach is that the individual fixed effects account for fixed, time-invariant differences across people, which includes many dimensions of ability and preferences that are unlikely to change over time. A main drawback of this approach is that one needs pre-collegiate earnings to be able to identify this model, as it essentially is estimated off of the within-person change in earnings surrounding degree receipt or enrollment. Hence, this approach can be used to study students who enroll later in life and who thus have substantial pre-collegiate earnings. Individual fixed effects models have been used extensively when studying the returns to vocational and sub-baccalaureate degrees as well as graduate degrees, since these programs tend to enroll older students with prior work experience.

To see the assumptions embedded in this approach, it is helpful to write out the formula for the treatment effect on the treated using our 3-period example. The treatment effect on the

¹⁹ This assumption often is discussed in two parts: 1) parallel pre-treatment relative trends in outcomes across treated and untreated units, and 2) no unobserved shocks that align with the timing of treatment and differentially affect treated units.

²⁰ See Card (2001) for a discussion of this issue in the context of schooling.

treated is calculated as the difference between period 3 earnings (post-degree) to period 1 earnings (pre-degree) among those who obtain the degree:

$$\begin{aligned}
TT_{ct3}^{FE} &= E[w_{icj3} - w_{i0j1} | C_{ci3}] \\
&= \sum_j \int_{A,Q} p_{c3}(j|A_3, Q_3) w_{cj3}(A_3) dF_3(A_3, Q_3 | C_{c3}) \\
&\quad - \sum_j \int_{A,Q} p_{c1}(j|A_1, Q_1) w_{0j1}(A_1) dF_1(A_1, Q_1 | C_{c1}) \tag{6}
\end{aligned}$$

As highlighted in Altonji and Zhong (2021), there are three reasons why earnings could differ across periods. First, ability and preferences could change, such that $dF_3 \neq dF_1$. Second, workers can gain experience in the job market that affects their occupational choices, and third, experience could affect occupation-specific earnings.

These different sources of earnings changes lead to three main identification assumptions. The first is that ability and preferences are stable over time, which can be a strong assumption. Note that the A and Q are defined to be net of c , so collegiate attainment cannot generate changes in these parameters.²¹ However, individual tastes for different professions or their productive capacity in different jobs may shift over time in ways that can bias this estimator. When this happens, period 1 (pre-collegiate) earnings are no longer an accurate counterfactual for period 3 earnings. For example, imagine a worker is a cashier at a retail store and finds that she would like to become a nurse. Her shifting preferences may induce her to get a degree in nursing and take a job as a nurse. But then the comparison of earnings from after versus before she obtained her degree will likely overstate the return to a nursing degree, as her change in preferences would likely induce her to shift towards a higher-paying medical profession even absent obtaining a nursing degree. To overcome problems associated with changes in preferences and ability, we need to observe pre-collegiate earnings *after* these preference changes have occurred.

A related problem is ‘‘Ashenfelter’s Dip’’ (Ashenfelter 1978). This phenomenon refers to the fact that those experiencing an adverse shock are more likely to enroll in a given program. In this context, workers who experience a job separation or an earnings decline may be more likely

²¹ More specifically, any changes to Q and A due to college enrollment is part of the return to enrolling in college.

to subsequently enroll in a degree program. This would cause a positive bias in the estimates, as the pre-enrollment earnings are artificially low.

The second identification assumption relates to growth in earnings as workers age and obtain experience. Earnings tend to rise with worker age, which can affect within-worker comparisons over time surrounding program enrollment. The main assumption one needs to invoke is that the choice of degree program is uncorrelated with the age profile of earnings. This can be violated, for example, if those who obtain certain degrees have higher returns to experience. This secular earnings growth will be attributed to the degree, causing an upward bias. To account for this problem, some researchers include individual-specific linear time trends in their individual fixed effects models (e.g., Stevens, Kurlaender, and Grosz, 2019). While these controls are helpful in addressing concerns related to secular earnings growth, they do not account for bias stemming from secular earnings increases that occur post-enrollment. It is, however, not clear where such variation would come from.

The third assumption embedded in the individual fixed effects model is that pre-collegiate occupational choices are separable from period 2 college investment decisions. If workers take jobs prior to college that help prepare them for their future degree program and that pay less because of this non-pecuniary benefit, the pre-collegiate earnings will be biased downward and thus the estimated return will be biased upward. This is a particularly important problem when using individual fixed effects models to estimate the return to graduate school. A student interested in law may work as a lower-paid paralegal, or a student interested in medicine may work for lower pay in a hospital or doctor's office to bolster their credentials for medical school. Ph.D. students often take lower-paid research jobs after college to prepare for graduate training. The strength of this assumption is likely to vary based on the specific set of degrees being studied.

The framework we presented above is useful in carefully assessing the assumptions of the main approaches researchers have taken to estimate the return to different types of postsecondary education. We now turn to a discussion of the empirical research. We will reference this framework and these assumptions throughout our discussion of these literatures in order to assess the validity of the identification assumptions invoked in specific circumstances.

4. The Return to Attending Different Institutions

This section focuses on the return to attending different types of institutions. As discussed in Section 2, there are many types of institutions in the U.S and around the world. Our discussion centers on the types of institutions and studies that receive the most attention: two-year colleges and four-year colleges. We begin by discussing conceptual issues and the context related to estimating the returns to two-year colleges, methods used, and the results from key papers. We then provide a similar discussion for research on the returns to four-year colleges. However, the four-year college literature is inseparable from the more general research question on the returns to college quality, and so the discussion is framed as such. We also discuss the emerging research on the returns to for-profit colleges in the U.S. The section concludes with a review of the international literature on college type and quality.

Prior to discussing the research, we stress several points. First, we note that there is a meaningful difference in assessing the return to beginning at an institution type and earning a credential from that institution. This is particularly important in the U.S., where college completion rates are low (Bound, Lovenheim, and Turner, 2010) and transfer is relatively common (Andrews, Li, and Lovenheim, 2014). Second, we frequently use the word “quality” and review the literature on “college quality.” Some researchers prefer the term selectivity or other variants because “quality” is not a well-defined term. In our discussion of the literature, we consider “quality” a theoretical construct and discuss how researchers use the different, imperfect, proxies available to them in the data. Third, given the large body of work on the returns to college type, we do not provide a complete literature review. Rather, we discuss the primary research questions and prominent examples of papers that address them, along with some outstanding issues that point to directions for future inquiry. We also focus on the methods, data, and identification strategies used in these papers that relate to Section 3. Our goal is to provide an overview of what is known and highlight promising avenues for future research through a better understanding of existing work.

4.1. Two-Year Colleges

Two-year colleges serve a wide range of students with a wide range of educational goals. The broad and varying nature of the student body implies that there is no single alternative educational option to serve as a counterfactual when assessing the differential returns to attending a two-year college. This is well established and at the center of the “democratization or diversion” analyses and debate (e.g., Rouse, 1995; Mountjoy, Forthcoming). The debate focuses on whether

two-year colleges generate increased degree attainment by drawing students into postsecondary education (democratization) or whether they reduce degree attainment by diverting students away from more resource-laden four-year colleges that have higher BA completion rates.

We begin by cataloging several broad ways in which researchers assess the returns to attending a two-year college. By doing so, we consider the value of attending a two-year college for students with different educational goals and different academic and demographic backgrounds.

In Section 4.1.1, we first consider students who are on the margin of no college versus a two-year college, or the so-called “democratization” margin. Research on the democratization margin includes many college non-completers, as two-year completion rates are quite low (Table 1). This has important implications for the research questions asked, the identification strategies employed, and the control groups that are used. Within the group of students on the democratization margin, we focus on students who begin their postsecondary schooling at two-year colleges with the goal of earning an associate or bachelor’s degree - those seeking a certificate, diploma, or just continuing education are discussed in Section 5. Degree-seeking students include part-time and full-time enrollees, who often are very different students with different completion rates.²² For example, in the high school class of 2004, almost 72 percent of full-time college enrollees earned a credential compared to 44 percent among part-time college enrollees.²³ Degree-seeking students at two-year colleges also include students who, after graduating high school, enroll directly in college and those who enter the labor market directly after high school, potentially for many years, or concurrently with postsecondary schooling.

When considering the return to two-year college versus no college degree, a key distinction is whether researchers are assessing the return to starting at a two-year college or the return to earning an associate degree. Both parameters are of interest, but they have different interpretations and lead to different policy implications. We thus discuss them separately below.

The second set of studies we discuss is on the returns to two-year college quality. This literature is sparse for a variety of reasons, including data limitations on measures of quality at the

²² Completion rates for full-time students tends to be much higher than for part-time students, but the literature comparing part-time and full-time students mostly relies on selection-on-observable approaches and is worthy of additional research using different methods.

²³ Author’s calculations using Education Longitudinal Study 2002. We exclude respondents who respond with “Equal mix of part-time and full-time.”

two-year level but also because there is no analogous notion of college choice as there is with four-year colleges. Most students outside major metropolitan areas only have a single public two-year college option, and the vast majority of two-year students attend their local public college. This feature of two-year colleges means that students do not express preferences over college quality attributes through their enrollment choices.

The third broad area of research we discuss in this section is the returns to attending a two-year college versus a four-year college, or the “diversion” margin. This discussion is often framed as the return to starting at a four-year college, but as we discuss in Section 4.2, this is often at the expense of starting at a two-year college. Two-year colleges on average have fewer resources than even open-access four-year schools. Hence, diverting students from four-year to two-year colleges could adversely impact collegiate attainment if these resources affect student outcomes.

Fourth, many students face the decision between enrolling in a public two-year college and a for-profit two-year institution. Analyses that include for-profits only appear in the literature on the return to for-profit colleges. The research broadly finds negative returns to enrolling in a for-profit college relative to a public college, and so the estimates are framed as a for-profit penalty. However, this can be reframed as the benefits of public two-year college relative to for-profit colleges. We discuss these considerations more in Section 4.3 in our discussion of the returns to for-profit colleges.

4.1.1. Return to First Enrolling in a Two-Year College Relative to No College

Enrolling in a two-year college versus no college has several potential costs and benefits. The most salient costs are the tuition and fees students pay to attend and the opportunity cost of foregone wages due to reduced labor supply while enrolled. There are two short-term and long-term benefits of enrollment. First, there is the option value associated with the enrollment choice. The main option value comes from the option to transfer to a four-year college and earn a bachelor’s degree, sometimes after completing an associate degree. Eighty-one percent of students who begin at a two-year college directly after graduating high school plan to earn a bachelor’s degree, despite only 14 percent earning that credential (Jenkins and Fink, 2017). This option value is highlighted in life-cycle models of educational investments and in structural models of college choice (e.g., Stange, 2012; Arcidiacono et al., 2016; Hendricks and Leukhina, 2017).

The second primary benefit to beginning at a two-year college for degree-seeking students is the potential earnings increase generated by accumulating college credits or earning an AA

degree. Only 17.9 percent of students who began at a two-year college in 2011 earned an AA degree within six years (Chen et al., 2019). While some went on to earn additional credentials or continue to pursue their degrees, about 44.6 percent were no longer enrolled, emphasizing the fact that any returns to starting at a two-year college is heavily determined by the return to some two-year college but no degree. Descriptively, seven years after graduating high school, students who begin their postsecondary education at a two-year college earn roughly 8-15 percent more per year than high school graduates who do not enroll in any college.²⁴ However, one cannot interpret these mean differences as the return to enrolling in a two-year college, since students with higher earnings capacity are likely to select into college.

Many papers seek to address this selection problem, often in a larger context of estimating the returns to partial schooling, certificates, and degrees (e.g., Rouse, 1995; Jepsen, Troske, and Coomes, 2014), using either a selection on observables approach or individual fixed effects with panel data (see Section 3 for a discussion of the assumptions underlying these methods). The research consistently finds positive impacts on earnings to attending a two-year college, largely driven by those earning credentials but also because of an earnings boost due to partially-completed schooling (e.g., Belfield and Bailey, 2017; Jacobson, LaLonde, and Sullivan, 2005).

Studies exploiting natural experiments on the returns to beginning at a two-year college for students on the margin between no college and college are few and far between. The increased prevalence of “free” two-year college and college promise programs (discussed in Section 6), where local residents pay no tuition to attend a two-year college, present opportunities to evaluate these marginal students. However, there are numerous difficulties in using these programs to assess the return to two-year colleges. First, diversionary incentives also compel students who would have otherwise attended a four-year college into a two-year college or vice versa (e.g., Carruthers, Fox, and Jepsen, 2020). Although imperfect, researchers have addressed this problem by creating subgroups of students most likely to be on each margin, using pre-college characteristics. Second, these programs have been implemented relatively recently, with little time to evaluate longer-run earnings outcomes. Third, the programs do not just influence college enrollment, they also influence financial outcomes and possibly peers’ decisions, violating any

²⁴ Authors’ calculations using the NCES’ Educational Longitudinal Study of 2002. The range is on the lower end when excluding zero wage earners and higher when they are included but consistent regardless of using average or median earnings.

instrument's exclusion restriction and possibly the SUTVA assumption that there are no spillovers from one treated observation to another.

Rouse (1995) uses distance to the nearest two- and four-year college as well as two-year and four-year in-state tuition and fees as instruments for college attendance. As discussed in Section 3, distance to the nearest college is a potentially problematic instrument if people who live closer to colleges otherwise have higher academic achievement and earnings potential. In Section 4.2.5, we provide a more detailed discussion on the validity of distance instruments in relation to four-year colleges. Similar identification issues arise when using distance as an instrument in the two-year setting. Rouse (1995) finds that two-year colleges increase overall educational attainment, suggesting a substantial democratization effect, however she also finds evidence that students diverted from four-year attendance complete fewer years of education.

Mountjoy (Forthcoming) advances the literature by using multiple instruments of a high schooler's distance to a two-year and four-year college to disentangle the diversionary and democratization margins of potential two-year enrollees. Using Texas administrative data on public K-12 and higher education students, he examines the role of distance to a two-year college holding distance to a four-year college constant (and vice versa) as well as accounting for neighborhood levels of urbanization and commuting zone fixed effects. This approach allows him to compare students within labor markets and who are in similar neighborhoods that differ in terms of their distance to the nearest college. Critically, he demonstrates that the distance instruments are conditionally uncorrelated with pre-collegiate student achievement measures (such as test scores), which supports his approach.

Mountjoy (Forthcoming) finds that for students on the margin of two-year college and no college, those induced to start at a two-year college "...complete 1.7 more years of schooling, are 26 percentage points more likely to earn a bachelor's degree, and earn \$1,337 dollars more per quarter around age 30 relative to never enrolling in college, which corresponds to a 21.8 percent earnings premium over their counterfactual potential outcome mean of \$6,141 dollars per quarter." However, he also finds that a third of new two-year college entrants are diverted from four-year schools and that these diverted students complete fewer years of education as a result. His results thus align with those in Rouse (1995). This estimation strategy has the potential to be replicated using other state's administrative data as well.

4.1.2. Return to Partially-Completed Two-Year College

The low completion rates as well as the high student loan default rate among two-year students (Looney and Yannelis, 2015) highlights the important question of whether there are returns to postsecondary credits among those who eventually drop out. The general consensus in the literature is that there is a return to partially-completed two-year college. For example, Kane and Rouse (1995) use NLS72 (and NLSY79) to show that students who enroll in a two-year college earn approximately 10 percent more than those who do not, even if they do not earn a degree. Restated in another way, they find that accumulating 30 credits (two semesters) is associated with a 4-6 percent increase in earnings. They employ selection on observables methods with rich pre-collegiate background and academic achievement controls. It is important to emphasize that the identification assumptions underlying these estimates are strong: even conditional on pre-collegiate achievement measures, it is likely that students who obtain fewer credits differ from those who obtain more on unobservables that correlate with attainment and earnings. Nonetheless, these are some of the best estimates to date on the return to credits among two-year dropouts.

While the estimates in Kane and Rouse (1995) rely on cohorts of students from decades ago, similar analyses using more recent nationally representative surveys from the high school graduating class of 1992 reveal strikingly similar results (Marcotte et al., 2005). The more recent surveys from the National Center for Education Statistics (Educational Longitudinal Study and High School Longitudinal Study) and NLSY97 provide continuing opportunities to re-evaluate the returns to partially completed two-year college with the same methods.

The series of papers that rely on state administrative data, as summarized in Belfield and Bailey (2017), provide further evidence on the returns to partially-completed two-year college. The authors of these studies use longitudinal datasets on quarterly earnings from state unemployment insurance systems, employing individual fixed effects models. They find positive returns to partially-completed schooling, even with as little as one semester (e.g., Jacobson, LaLonde, and Sullivan, 2005) or just a few credits. For example, Liu, Belfield, and Trimble (2015) show that men (women) in North Carolina who earned 1-10 credits earn \$1,162 (\$638) more than those who enrolled but earned zero credits after either dropping out or failing their course(s). Beyond the first few credits, the returns to additional credits is roughly linear (Belfield and Bailey, 2017). While individual fixed effects allow researchers to account for unobserved fixed heterogeneity across students, these papers necessarily focus on older students who have sufficient pre-collegiate earnings.

The finding that there is a positive return to sub-degree credits in two-year colleges is not universal. Kreisman, Smith, and Arifin (Forthcoming) show that one-third of college non-completers, especially from two-year colleges, who post their resumes on an online jobs board do not list that college on their resume. The fact that the job applicants are willing to hide that information demonstrates that they do not believe there is value in the enrollment spell or at the very least, that the negative signal from dropping out outweighs any benefits to demonstrated human capital accumulation at the job application stage. It is possible the eventual wages of these job applicants reflect the human capital attained in the partially-completed schooling, but no such evidence yet exists.

Additionally, the recent growth in resume audit studies sheds light on the value of partially-completed two-year colleges. In the resume audit literature, researchers send fictitious resumes to real job postings and track the callback rate of the job application. The experimental design relies on randomly assigning resume attributes, including colleges enrolled, fields of study, and/or degrees earned. Darolia et al. (2015) find that resumes listing two-year colleges but no degree have positive but not statistically significant effects on employer responses and interview requests. Although the randomized worker attributes provide clear identification in these studies, the relationship between receiving a callback and ultimate employment and wages is unclear and largely unanswerable in the resume audit literature (Heckman, 1998). The returns to schooling may depend on how actual job seekers conduct their search before and after a target application to a firm and how firms set their wages.

4.1.3. Returns to an Associate Degree

Another strand of literature examines the return to an AA degree relative to no college. Practically speaking, this is not commonly assessed because a natural control group to AA degree earners are students who enroll in a two-year college but do not earn a degree. AA degree completion (among those who enroll) is about 18 percent (Chen et al., 2019), so there are plenty of potential students for a control group with common support on observed characteristics to the degree earners. Still, we also must be concerned about how degree completers and non-completers differ on unobserved characteristics.

In commonly-used, publicly available datasets, such as Current Population Survey or American Community Survey, there are typically broad categorical variables on educational attainment, including degrees earned and whether there is “some college.” Unfortunately, this

means that when assessing the value of an AA degree, the comparison group is comprised of a wide range of different types of students. Non-degree holders may include people who earned a high school diploma and did not enroll in college, those who enroll in two-year colleges and did not earn a degree, and those who enrolled in four-year colleges and did not earn a degree. The people in these groups are very different from one another, which highlights the need for more detailed educational data than those readily available in these datasets.

Given the prevalence of degree non-completion, the literature largely focuses on the return to an AA degree compared to some college but no college degree. The ideal experiment to identify this parameter is to take two students with the same amount and type of schooling but to randomly assign one of the two people a two-year degree. Perhaps the best approximation to this experiment is the “sheepskin effect” literature, whereby researchers look for a discontinuous jump in earnings among two groups of people with about the same amount of schooling, but only one of whom earns the degree.²⁵ The focus of this research thus is on identifying whether there is a return to obtaining a degree over and above the credits one earns for that degree. Some studies examine years of schooling to see if there are larger earnings increases for years in which degrees are earned (e.g., Hungerford and Solon, 1987; Jaeger and Page, 1997), while others rely on college credits earned coupled with information on degree receipt (e.g., Kane and Rouse, 1995).²⁶

Kane and Rouse (1995) and Jaeger and Page (1997) set the stage for modern analyses of sheepskin effects using nationally-representative datasets. Despite their different contexts, they both find evidence of substantial sheepskin effects of an AA degree. Kane and Rouse (1995) use the National Longitudinal Survey of the High School Class of 1972 (NLS72) and show that women who obtained an AA degree earned over 25 percent more than those who did not obtain a degree, conditional on number of college credits (effects for men are only present for a BA degree). They find similar results for women in the National Longitudinal Survey of Youth (NLSY1979) as well as similar-sized sheepskin estimates for men and women in these data. Using the 1991 and 1992 March CPS, Jaeger and Page (1997) find a 7.5 percent increase in hourly wages for occupational AA degrees and an over 20 percent increase in hourly wages for an academic AA degree. Their

²⁵ Sheepskin effects are so named because college diplomas historically were printed on sheepskin.

²⁶ Jaeger and Page (1997) show that using “true” measures of degree receipt removes biases from assuming degrees are earned based on a certain number of years of schooling (e.g., 12, 16).

documented effect is for White men and women, while they cannot rule out a null effect for Black men or women.²⁷

There are a few notable critiques of the sheepskin effect literature that examines earnings boosts after flexibly controlling for years of schooling or accumulated college credits. The first is that estimates are sensitive to the choice of functional form and separately to misreporting or measurement error of schooling and degrees (Flores-Lagunes and Light, 2004). Second, there are inherent and likely unobserved differences between people with similar amounts of schooling or credits but who differ with respect to degree status. These unobserved differences could be correlated with earnings and generate bias in the estimates on the return to a two-year (or four-year) degree. For example, it could be that conditional on having sufficient credits to graduate, the student who gets the degree is better able to navigate degree requirements and rules. Alternatively, students who drop out right before degree completion could experience a negative shock (such as a sick parent) that adversely impacts subsequent labor market outcomes as well. Advances in data availability presents opportunities to control for some of those unobservables, such as courses taken and performance in those course, sequences of courses taken, and the accumulation process of credits. However, it is unlikely that even a very rich dataset would allow one to account for all of the factors that drive degree completion conditional on college credits.

Availability of state administrative data linking schooling to earnings data has made it possible for researchers to move beyond sheepskin effect analyses in order to assess the return to an AA degree relative to not enrolling in college. The administrative data capture precisely-measured credits and degrees earned. Perhaps most importantly, and distinct from nationally representative surveys, they are longitudinal in nature. The longitudinal data allow researchers to observe the evolution of earnings and in many cases, earnings prior to schooling. This novelty compels almost all researchers with these data to employ an individual fixed effects regression to assess the return to an AA degree.

Table 4, adapted from Belfield and Bailey (2017), summarizes the findings from the individual fixed effects analyses using state longitudinal databases to estimate the return to an AA degree. Across eight different states, researchers find positive effects of earning an AA degree on quarterly earnings 5-9 years after college entry. The average gain across states for men is \$1,160

²⁷ The point estimates for Black men are roughly the same size as for White men but the standard errors are three times as large. However, point estimates for Black women are negative.

and for women is \$1,790. As their abstract summarizes, these estimates imply that “completing an associate degree yields on average approximately \$4,640–\$7,160 per annum in extra earnings compared to entering college but not completing an award.” However, as discussed in Section 3, the fixed effects regression with state longitudinal data systems are not without their own identification challenges. Additionally, the estimates are identified from people with earnings histories prior to college enrollment, which limits the generalizability of the results.

Along with the growing availability of data, there are a number of resume audit studies, some of which assess the value of an AA degree. Darolia et al. (2015), Deming et al. (2016), and Deterding and Pedulla (2016) all run resume audit experiments and find little evidence that listing an AA degree has a meaningful impact on employer callback rates relative to no college or a high school diploma.²⁸ As previously discussed, though these resume audit studies are well identified and show little benefit to callbacks from an AA degree, they do not necessarily imply that the labor market return to an AA degree is zero.

4.1.4. Two-Year College Quality

Two-year colleges across the U.S. have varying completion rates, transfer rates, and average earning among enrollees. These differences highlight the importance of understanding why some two-year colleges have better outcomes than others and whether this is driven by differences in students or the quality of the education offered. However, researchers rarely assess the returns to two-year college quality. There are two main reasons for the lack of research in this area. First, many of the variables used to assess four-year college quality (discussed in Section 4.2) are not relevant or are not available for two-year colleges. Measures like applications, admission rates, and yield rates are not relevant to and/or do not vary across these open-access institutions. Additionally, the ACT or SAT scores of enrollees are unavailable to researchers because a large fraction of two-year students either do not take the exams or they take the exams but do not submit them to colleges. Second, most students who want to attend two-year colleges only have a single viable two-year college option. These students and decisions often are driven by preferences for location coupled with financial considerations at these lower-tuition public institutions (Long, 2004). This implies that students are unlikely to select two-year colleges based

²⁸ Many of the job postings in Deming et al. (2016) required or preferred a bachelor degree and so a callback is not necessarily expected for applicants with a high school diploma or an associate degree.

on quality factors because of the local nature of enrollment. To date, only a few papers examine the effect of two-year college quality.

Stange (2012) provides some of the clearest evidence on the returns to two-year college quality. He uses the fact that students attend the local two-year college, which mitigates selection concerns considerably. One's local two-year options define the college quality to which they have access, and he examines whether postsecondary attainment varies with measures of two-year college quality, such as expenditure per student. He finds that these quality measures have no impact on educational attainment. Although the analysis is not linked to labor market data, the previously-discussed positive returns to two-year degrees suggests that there are few labor market returns to two-year college quality (as he measures it). That is not to say there is no return to two-year college enrollment, but the variation in quality exploited by Stange (2012) does not reveal meaningful educational attainment impacts.

Smith and Stange (2016) overcome one data limitation of two-year colleges by using administrative data from the College Board linked to college enrollment data from the National Student Clearinghouse (NSC).²⁹ They calculate the average PSAT scores of enrollees and use it as a measure of college quality. The PSAT is widely taken across the U.S., even among students who attend two-year colleges. Using average PSAT scores of enrolled students as a measure of college quality has a similar interpretation to the research that relies on the average SAT scores of enrollees as a measure of quality in the four-year literature. They find substantial variation in average PSAT scores of enrollees across two-year colleges, even within the same state.

The authors then assess the impact of college quality on AA and BA degree attainment from the NSC, within and across two-year and four-year colleges. They control for demographics, individual PSAT score, and even high school attended in their selection on observables approach. The findings indicate that a one standard deviation increase in college quality increases the BA attainment rate for students who begin at a two-year college by about 3 percentage points, which is meaningful but much smaller than the relationship between peers' PSAT scores and degree attainment for students who begin at four-year colleges. However, the identification strategy relies on a strong set of assumptions as to whether the students at two-year colleges of varying quality

²⁹ NSC data are a near census of college enrollment spells and credentials earned in the U.S. For-profit college enrollment is the biggest data coverage deficiency, but it includes some of the largest for-profits in the U.S. See Dynarski, Hemelt, and Hyman (2015) for additional details and practical considerations for researchers.

have the same ability and preferences. Their controls for pre-collegiate academic performance through PSAT and SAT certainly reduce bias related to ability, but there are no explicit controls for preferences beyond basic demographics and high school attended. Overall, conceptual difficulties of defining two-year college quality, data deficiencies, and strong identifying assumptions used in this literature point to a promising area of future research, especially as new data sources become available.

4.2. Four-Year Colleges and College Quality

In this section, we review the research on the returns to four-year colleges and college quality more generally. As discussed in Section 2, four-year colleges differ from two year-colleges in the students they serve, admission processes, tuition, educational and non-educational expenditures, degrees awarded, and critical to this section, the range of quality of institutions.

We begin by discussing the conceptual issues involved with estimating the returns to four-year college, which revolve around the counterfactual colleges that four-year enrollees consider. Students at the most-selective four-year colleges typically consider a range of four-year college qualities. Many students also consider both two-year and four-year colleges, so the return to four-year colleges is related to the two-year and four-year college tradeoff. Additionally, we briefly discuss the return to partially-completed four-year college and BA degrees.³⁰

Next, we give an overview of the methods used by researchers before going into detail about the findings from each method. Some of the methods are connected to colleges' policies and procedures (e.g., studies employing regression discontinuities and minimum admission criteria), and so they are discussed in tandem.

4.2.1. Conceptual Issues

There are several broad issues in estimating the returns to a four-year college. First, there is the return to starting at a particular four-year college, which requires an understanding of the unobserved counterfactual enrollment. The alternative to starting at a particular four-year college includes a two-year college, no college, or a different four-year college. Determining the types of colleges students consider *ex ante* is challenging and often impossible. Several datasets and studies in the U.S. provide information on college applications to four-year colleges, which provide a source of information on four-year colleges considered but, as far as we are aware, there are no ordered preferences of applications to better understand the counterfactual enrollment. Ranked

³⁰ Bound and Turner (2011) provide an in-depth review of college completion and the associated literature.

preferences of colleges are common outside the U.S., which we discuss in Section 4.4. Additionally, four-year college students may enroll in or consider open access institutions at the two-year or four-year level that do not require applications. There are also many four-year enrollees at the “diversion” margin, who considered two-year colleges (and vice versa), especially at the least-selective four-year colleges. Some four-year enrollees additionally consider no college or at least substantially delayed enrollment. This is a relatively small fraction of four-year students compared to those who considered some type of college.

Second, researchers are interested in estimating the returns to both four-year college quality and college quality more generally. Putting aside the definition of quality, there are few studies that distinguish the returns to four-year college quality from college quality in general. The fact that so many students consider both two-year and four-year colleges explains researchers’ focus. The studies that exclusively examine four-year college quality either only examine elite colleges where most students did not consider a two-year college or do not fully capture the counterfactual options in their estimation procedure.

Third, there is no consensus on the definition of “college quality,” and in fact some researchers avoid the term and instead use terms such as “selectivity.” Perhaps one of the most common measures of college quality used in the literature is the average SAT scores of enrollees.³¹ However, most papers recognize that the single measure is imperfect, so they estimate models using alternative measures of quality. Those measures frequently include expenditures per student, number of applications, admit rate, yield rate, and persistence and completion rates. More often than not, approaches using these various measures find consistent results. One value of expenditures per student and completion rates is that they are available for almost all colleges, including two-year colleges. As Section 2 describes, on average, two-year colleges have substantially lower expenditures per student and completion rates than do four-year colleges (see Tables 1 and 2). Because of these differences, it is common for researchers to compare the returns to different postsecondary levels as another way of assessing the returns to college quality.

Black and Smith (2006) show that any single measure introduces substantial measurement error that can bias the estimates. For example, they report that using SAT scores alone attenuates the estimated return to college quality by up to 20 percent. They propose a variety of alternatives,

³¹ Technically, most researchers take the average of the 25th and 75th percentiles of the math and verbal SAT scores of enrollees, as these are the measures available in IPEDS.

including a factor analysis that combines all their available measures of quality into a single index. As an alternative to specific college characteristics, some researchers make use of external sources of quality, including *U.S. News and World Report's* College Rankings and Barron's Selectivity Categorization (e.g., Brewer, Eide, and Ehrenberg, 1999; Bound, Lovenheim and Turner, 2010, 2012; Sjoquist and Winters, 2016). The *U.S. News and World Report* rankings takes into account many of the aforementioned measures but assigns a subjective weight to them. These rankings are not comprehensive across all institutions, and so they tend to be used by researchers when considering quality at elite institutions. Barron's is a comprehensive measure for all institutions, but it is categorical so there is less variation to exploit than with continuous measures. Some researchers also focus on flagship public colleges, which are the most selective college in each state, as an indirect measure of quality (e.g., Hoekstra, 2009; Andrews, Li, and Lovenheim, 2016).

Fourth, while there is value in estimating the return to a BA degree, *per se*, there are few rigorous studies in this area compared to the analogous setting of two-year colleges and AA degrees. One explanation for the lack of research on this question is that there are few exogenous sources of variation that impact BA degree attainment but not labor market or educational attainment outcomes. Additionally, the weaker assumptions in the individual fixed effects models used to assess the returns to an AA degree are not feasible in the four-year setting, because so few students have prior work experience.

Fifth, it is important to understand whether there are returns to partially-completed four-year college. But once again, there are relatively few studies that examine this question. One reason for the lack of work in this area is that completion rates at four-year colleges, which are just above 60 percent within six years of initial enrollment, are higher than two-year colleges. Another reason is that there are few identification strategies that can address selection into graduation.³² Kane and Rouse (1995) find no evidence that additional credits at four-year colleges among eventual dropouts increase wages using the sheepskin effect approach. But the selection on observables identification strategy requires strong assumptions. Additionally, their analysis uses data from NLS72, and effects could differ with more recent cohorts.

³² Ost, Pan, and Webber (2018) assess the returns to college persistence for students with college GPAs around a dismissal threshold in Ohio. Using an RD, they find that those just below the threshold are 10 percentage points less likely to earn a BA and earn 4.8 percent less than those above the threshold. This cannot be interpreted as the return to partially completed schooling relative to a BA, but the degree completion effect likely contributes to the earnings estimate.

Sixth, some research considers the academic “match” of students with a college. Academic match is a measure of how closely students’ academic ability aligns with that of college enrollees at the institution of choice. The general idea is that when assessing the returns to a college or college quality, there are portions attributed to the student, to the college, and potentially to the match over and above the other two contributing factors. We discuss the research on student-college match in more detail in Section 4.2.6, including the concept, measurement, and the estimated effects of match.

4.2.2. Overview of Results and Methods

Descriptively, there are substantial differences in the average earnings across the four-year college quality distribution. Students who begin at “more selective” four-year colleges earn about 20-30 percent more per year than students who begin at “inclusive” four-year colleges.³³ Additionally, seven years after graduating high school, students who begin their postsecondary education at a four-year college earn roughly 23-35 percent more per year than high school graduates who begin at a two-year college.³⁴ The research in this space seeks to estimate what fraction of those earnings differences can be attributed to the institution itself and its quality.

Table 5 highlights some of the findings from the literature on four-year college quality and college quality in general in the U.S. With few exceptions, college quality positively impacts college completion and earnings, particularly when the counterfactual includes two-year colleges. The table is not exhaustive but shows representative results in a variety of contexts and using a variety of methods and identification strategies.

There are, broadly speaking, two identification strategies that map to Section 3. The first is the selection on observables approach, which is chronologically where the literature begins and the method that requires the strongest assumptions. However, data advances have allowed researchers to expand the set of observables (e.g., including high school test scores) to remove some, but not all, selection concerns. We discuss this research as well as the extension of this method to include comparisons of siblings/twins and college application portfolio controls, the latter of which is only relevant to four-year colleges.

³³ Authors’ calculations using the NCES’ Educational Longitudinal Study of 2002 and vary depending on whether we exclude zero wage earners or consider mean or median. “More selective” and “inclusive” are Carnegie Classifications.

³⁴ Authors’ calculations using the NCES’ Educational Longitudinal Study of 2002. The range is on the lower end when excluding zero wage earners and higher when they are included but consistent regardless of whether we use average or median earnings.

Next, some researchers use instruments to identify the causal impact of college quality. Recall, these approaches assume that conditional on the instrument, the distribution of ability and preferences are the same on average between those who enroll in a high quality and low quality college. In practice, the papers that rely on instruments fall into two further categories: regression discontinuities and other methods. The RD papers exploit discontinuities in college admissions criteria, such as SAT score or high school GPA. The non-RD papers rely on variants of distance instruments or policy rules and changes that lend themselves to a difference-in-difference approach. We discuss the specific papers, methods, identifying assumptions, and results of the returns to college quality in sequence of the approaches described.

4.2.3. Selection on Observables

Studies that rely on a selection on observables approach tend to find large positive effects of the impact of college quality on college completion and earnings. But the data used influences the set of observables and consequently, the identifying variation. This leads to variability in the strength of the identifying assumptions across studies.

The first broad research question that relies on a selection on observables approach is what are the returns to college quality. Brewer, Eide, and Ehrenberg (1999) is an early and leading example of this strand of the literature. The authors ask whether attending higher quality colleges increases earnings at various points after college. They use the nationally representative samples NLS72 and High School and Beyond, which collectively include the high school cohorts 1972, 1980, and 1982. An appealing feature of these nationally representative samples, particularly in contrast to administrative data, is the rich set of background variables to control for selection into different colleges. The authors control for parental income and education, high school GPA, and college entrance exam scores, along with more traditional background variables. They show that these primary variables strongly predict the type of college students enroll in and accordingly, they should be used in almost all selection on observable studies. However, the relatively parsimonious number of control variables in their model leaves open the possibility that there is a missing variable correlated with college quality and earnings, perhaps measures of family wealth, student motivation, or occupation preferences. Nevertheless, they find that attending a top private college, as determined by Barron's selectivity, leads to 26-30 percent higher earnings relative to public four-year college in the least selective group. Findings consistent with these results have been

found in a series of other papers that use different nationally representative samples and slightly different methods (e.g., Long, 2008; 2010).

There are two main ways in which the approach and findings in Brewer, Eide, and Ehrenberg (1999) have been extended by subsequent research. First, Andrews, Li, and Lovenheim (2016) assess a similar question but use administrative data from Texas. One of the values of administrative data relative to nationally representative data is that there are many students at each college, and so researchers can assess the returns to particular colleges. However, they have to focus on people who stay in Texas from high school into a public college and then into the labor market. In this paper, the authors examine the return to graduating from UT-Austin or Texas A&M relative to a non-flagship public college.³⁵ Andrews, Li, and Lovenheim also depart from the literature by estimating quantile treatment effects as opposed to mean effects.³⁶

They find that there is a positive and meaningful economic return to starting at a flagship college relative to non-flagship public colleges, but there are substantial amounts of heterogeneity in the returns. The earnings premiums increase across the earnings distribution, from a low of 2.7 percent at the 9th percentile to a high of 31.7 percent at the 97th percentile at UT-Austin. There is less heterogeneity among Texas A&M graduates. These findings highlight that individual colleges with similar broad categorizations of quality or selectivity can impact students differently from one another. Additionally, colleges, both within a single college or across a set of colleges, have a distribution of effects on its students. Mean earnings returns thus may come with varying amount of *ex-ante* risk, which is picked up by examining effects on the entire distribution of earnings. Field of study differences across colleges could explain some of this variation. The approach taken by Andrews, Li, and Lovenheim (2016) has the potential to be replicated across most settings that use a selection on observables approach.

Second, Black and Smith (2004) extend the selection on observables approach using matching techniques. They investigate the returns to college quality with a nationally-representative dataset (NLSY79), using propensity score matching methods and comparing their results to regression-based methods that are more frequently used. One of the main points the

³⁵ They also consider the returns to starting at these colleges, not necessarily graduating. However, their results are similar when focusing on a sample of graduates.

³⁶ They follow the unconditional quantile treatment effects methods outlined in DiNardo, Fortin and Lemieux (1996) and Firpo (2007).

authors highlight is that matching methods allow researchers to assess the common support of students across the college quality distribution. Examining common support addresses the concern of whether there are students of similar backgrounds and abilities at both high and low quality colleges. The authors find that the common support is only “weakly satisfied,” which also implies that the functional form in a linear regression may be an important determinant of the estimated coefficients on the returns to college quality. In relation to the model in Section 3, a lack of common support necessarily implies that there do not exist students with the same ability and preferences at two different college types, and therefore one cannot estimate the treatment effect of college type or quality. Black and Smith (2006) find that matching methods and OLS yield similar estimates on the return to college quality for men but not so for women. Perhaps more importantly, they highlight the assumptions required to obtain causal treatment effects using a selection on observables approach and suggest that researchers consider issues such as common support to strengthen the validity of their results.

The second broad research question that relies on a selection on observables approach is what are the returns to beginning at a four-year college over a two-year college. The question is often assessed from the opposite side of the same coin and framed as the two-year college “penalty.” The penalty on the probability of earning a BA degree can be as high as 30 percentage points (Smith and Stange, 2016), and the penalty on earnings can be between 5 and 16 percent, depending on the context, specification, and gender of the student (e.g., Reynolds, 2012).

Reynolds (2012) is a leading example of the selection on observable approach to answer this question.³⁷ Reynolds uses the National Education Longitudinal Study (NELS:88) that follows approximately 12,000 eighth graders in 1988 through college and into the labor market. These data contain a wide range of variables that allow him to control for many dimensions of selection into two-year and four-year colleges. One unique and important variable missing in previous analyses that Reynolds (2012) is able to include is students’ desire to complete a BA, even among those who begin at a two-year college. This helps, to some extent, control for a potential source of bias in preferences over degrees, majors, and occupations.

Using both OLS and matching methods, Reynolds finds that starting at a two-year college over a four-year college reduces the probability of earning a BA by 23 and 25 percentage points for men and women, respectively, even among BA-aspiring students. He also finds a reduction in

³⁷ Numerous other papers find similar two-year penalties (e.g., Rouse, 1995, 1998; Leigh and Gill, 2003).

earnings between about 5 and 10 percent, especially for women. The estimates on earnings are a bit imprecise and smaller in magnitude than some other studies. Lastly, Reynolds (2012) assesses the extent of selection and whether it could explain his estimates. He uses a Mantel–Haenszel nonparametric test statistic and determines that his results are unlikely to be entirely driven by selection. This is one of the few papers on returns to college quality that assesses the selection on observables approach with this method. Since the identifying assumptions for selection on observables approaches laid out in Section 3 are difficult to verify and also are strong, the literature that uses similar methods would be strengthened by employing similar sensitivity checks. Oster (2019) offers a sensible approach that uses the movement in the coefficient and R-squared as researchers add increasingly more controls in their model to assess the likelihood of residual bias in selection on observables studies. This approach could easily be employed in studies of the return to college quality.

There are two additional variants to the selection on observables approach that appear in the literature on the returns to college quality: the use of twins or siblings and college application portfolios. Data on twins, and more generally for families and siblings, are increasingly available within administrative data. Researchers have used these data to include family fixed effects that serve as controls to account for selection issues associated with unobserved and time-invariant variables (e.g., Bhuller, Mogstad, and Salvanes, 2017; Chetty and Hendren, 2018). These types of family-related unobservables are perfect examples of factors likely to be correlated with college choice and earnings that violate the identifying assumptions in the selection on observables models. However, as Bound and Solon (1999) discuss in detail, twin and sibling fixed effects are not without problems. Siblings and twins are still different from one another in ways that may be correlated with college choice and earnings.³⁸

Behrman, Rosenzweig, and Taubman (1996) is the first paper to use twins to assess the returns to college quality and type.³⁹ They surveyed 941 pairs of twin sisters born in Minnesota between 1936-1955, sourced from the Minnesota Twin Registry. Their findings show the importance of using the twin fixed effect models over OLS and that there are positive earnings returns to certain college characteristics, including Ph.D. granting institutions, private colleges,

³⁸ Bound and Solon (1990) make this point and emphasize that adding in a twin fixed effect removes some of the exogenous sorting into college that was available in the cross-section, which can induce bias in the results.

³⁹ Several prior papers use twins to assess the returns to number of years of schooling (e.g., Ashenfelter and Krueger, 1994). In practice, college quality impacts years of schooling (Zimmerman, 2014).

well-paid senior faculty, and smaller enrollments. Though these variables differ from standard measure of college quality used in the literature, the paper demonstrates that the selection on observables approach benefits from accounting for selection bias related to family-level unobservables.

Dale and Krueger (2002) were the first to use college application portfolios, which is the second variant on the selection on observables approach that we discuss. The authors compare the college selectivity and earnings of students who apply to and are accepted to a similar set of colleges. By controlling for the set of applications, the authors make strides in removing a potential source of bias in students' abilities and preferences that may be correlated with the college quality to which students enroll. By controlling for where students are accepted, the authors argue that they control for unobservable ability, motivation, and interest in each college, because the admission committee has access to information that the econometrician does not. It is analogous to the impossible task of controlling for all the information in a college application.

This estimation strategy helps mitigate selection issues as to why students choose different colleges. However, the main critique of this approach is that even conditional on applications and acceptances, there likely are remaining selection issues that compelled students to choose to enroll in different colleges from one another. Furthermore, there is concern that the quality differences between colleges in students' effective choice sets are small. In relation to Section 3, it is likely that the approach reduces selection bias, at least based on academic ability. However, it is less clear whether the strategy removes selection bias induced by preferences or other dimensions of college characteristics, since the choice of different colleges reveals some degree of preference heterogeneity.

Using *College and Beyond*, a survey of students from 34 elite colleges, Dale and Krueger (2002) is one of the few papers in the literature to find null effects of college quality on earnings. They do find an eight percent increase in earnings for students from low-income families though. Dale and Krueger (2014) update their findings with administrative earnings data and find similar null results. Ge, Isaac, and Miller (Forthcoming) revisit the dataset with the same methodology but do not limit the sample to full-time full-year workers. The slightly different sample obtains the

same results for men, but women who attend a college where the average SAT score of enrollees is 100 points higher experience 14 percent higher earnings.⁴⁰

It is important to note that the College and Beyond data are from the most selective colleges in the U.S., so the results in Dale and Krueger (2002; 2014) do not speak to the broader issue of college quality, such as comparing non-selective to selective colleges or the two-year/four-year tradeoff. Long (2008) investigates the role of college quality across the entire range of four-year college quality, as measured by several college attributes. He partially confirms Dale and Krueger's results across the full range of four-year college quality using NELS:88 but finds that the results are sensitive to the measure of college quality. The results also are imprecise, so he cannot rule out meaningful positive effects of college quality on BA completion or earnings.⁴¹ Smith (2013) also considers all four-year colleges in the U.S., but finds different results than those in Long (2008) and Dale and Krueger (2002; 2014). Smith controls for application portfolios using several million students who take the SAT in the mid-2000s. He finds that a 100 SAT point increase in the average SAT of college enrollees corresponds to a 5 percentage point increase in the probability of receiving a BA. Additionally, he estimates twin fixed effect models with over 11,000 sets of twins, and the results are almost entirely unchanged. This provides some validation to the Dale and Krueger method by ruling out selection issues related to family background, however the findings differ across the two studies.

Mountjoy and Hickman (2021) estimate the “value-added” of public colleges in Texas through state administrative data via a Dale and Krueger (2002) approach. A primary distinction with Dale and Krueger (2002) is that they calculate the value-added of each college as opposed to estimating returns to college quality based on a proxy for college quality (e.g., average SAT of enrollees). Conceptually, value-added is similar to what other researchers try to estimate – the effect of the postsecondary institution on educational and labor market outcomes, net of the contribution of the student who attends the institution to those outcomes. The state administrative data allow them to separately estimate the return to each public college in the state. Interestingly,

⁴⁰ They also find that women are 5 percentage points more likely to earn an advanced degree and 4 percentage points less likely to get married. There are no such effects on degrees or marriage for men.

⁴¹ Long (2008) goes a step further and compares a variety of methods, including OLS, controlling for the application portfolio, a distance-related instrumental variable, and Black and Smith's (2006) multiple proxy method. Although there is some sensitivity to the estimates, the variability depends on whether the outcome is degree attainment or earnings. Better understanding the extent of bias across methods has the value of allowing the use of these imperfect methods while acknowledging potential weaknesses.

they show that their value-added estimates are only weakly related to existing measures of selectivity or quality. This novel finding suggests that the notion of quality is not as simple as the typical measures available to researchers, partially due to the role of student selection into colleges. Access to administrative data have and will continue to increase the number of studies that report college-specific results that do not rely on aggregate measures of quality (e.g., Kirkeboen, Leuven, and Mogstad, 2016; Chetty et al., 2017).⁴² Overall, Mountjoy and Hickman (2021) find that attending a college with a one standard deviation higher value-added increases the probability of completing a BA by 3.7 percentage points and increases subsequent earnings by 3 percent - both relatively modest effects that are much smaller than unconditional differences across colleges.

4.2.4. Regression Discontinuities and Minimum Admission Criteria

There are a series of papers that use regression discontinuity designs to estimate the returns to college quality and type on college completion and labor market outcomes. In the U.S., all of the papers exploit a discontinuity in the probability of admission based on a minimum admission criterion, either high school GPA or SAT score. We detail the analogous studies in international contexts in Section 4.5. This general approach is a very powerful way to address selection issues, but the demanding data requirements coupled with the need for an admissions discontinuity make the papers few and far between.

Minimum admission criteria help to provide causal estimates on the return to college type or quality, but the existence of the criteria alone are not sufficient. First, similar to all RDs, the admission criteria need to have an impact on the probability of enrollment to a particular college, set of colleges, or college type. Second, to assess the impact of college type or quality, the counterfactual enrollment must be of a different type or quality than the college on which the study focuses. Third, there can be no manipulation of admission criteria to ensure that the threshold generates exogenous variation in college access and enrollment. There are some details particular to postsecondary education that can make this a strong assumption. For example, some admission criteria are publicly known to students and others are only known by the admissions committee (and somehow known by the researcher). Georgia's public universities have a publicly available statewide minimum SAT score. Conversely, Hoekstra (2009) makes use of an unpublished SAT

⁴² It is likely that several state administrative data systems allow for a similar approach to Mountjoy and Hickman (2021). Broecke (2012) uses a Dale and Krueger (2002) type approach in the U.K. but has the additional virtue of observing student preferences over colleges to which they apply.

minimum score at a state's flagship university. Publicly known criteria are conducive to manipulation. Hidden criteria have been ascertained through institutional knowledge of a college's admission policies and also through statistical techniques that search for admission or enrollment discontinuities that can only be explained by such admission criteria. An example of the data demanding statistical techniques can be seen in Goodman et al., (2019), who find 21 colleges that use minimum SAT thresholds for admission but do not make their policy public. There are likely many more that researchers can explore.

Despite the appeal that hidden criteria reduce the potential for manipulation, publicly known criteria are still exploited in the literature but require extra consideration. Perhaps most importantly, when the criterion is the ACT or SAT, researchers make use of the first attempt's score. Students can and do retake college entrance exams, and Goodman, Hurwitz, and Smith (2017) show that they retake the SAT to get access to Georgia public universities (i.e., they exhibit manipulation). The authors argue that targeting an ACT or SAT score just above a known admission criterion is nearly impossible based on how the exams are scored. An unfortunate drawback of using the first ACT or SAT score is that the first stage becomes weaker.

High school GPA introduces some different challenges than college entrance exams, especially surrounding specific institutional contexts of how GPA is calculated. For example, Zimmerman (2014) discusses how public colleges in Florida employ different GPA calculations that are not made public and that do not correspond to the GPA students would see on their own high school transcript. Kozakowski (2020) shows graphical and statistical evidence of manipulation among high-income students in Massachusetts who need a minimum GPA to enroll in an in-state public college.

All the studies that use RDs on admission criteria find positive returns to college quality, despite focusing on very different types of college quality and potential enrollees. Hoekstra (2009) uses a minimum SAT score for admission to a flagship state university. The minimum SAT is not publicly known and so he uses the admission relevant SAT score, which is subject to retake. Because he uses the admission-relevant score, he finds a graphically compelling first stage – scoring above the SAT minimum increases the probability of enrolling in the flagship university by 37 percentage points. After scaling the reduced-form estimates by the first stage, he then finds that attending the flagship university increases earnings for White men aged 28-33 by 20 percent relative to attending the counterfactual institution. As the author acknowledges, the biggest

drawback of the paper is an inability to know where rejected applicants enroll and whether they remain in the state. He uses some additional sources that suggest the rejected applicants are likely to stay in state, and there are numerous other four-year colleges that are marginally less selective. But, unlike the next set of papers, the counterfactual distribution of colleges is unknown.

Zimmerman (2014) uses a high school GPA admission cutoff at Florida International University (FIU) – a relatively non-selective four-year public college. There are two notable differences with Hoekstra (2009). First, FIU attracts applicants and enrollees with lower academic scores than the selective flagship in Hoekstra (2009). Second, Zimmerman observes the counterfactual enrollment of students who do not enroll in FIU. Although some students just barely admitted forgo attending private or out-of-state four-year colleges, they are much less likely to attend two-year colleges. Men are 26.1 percentage points less likely to enroll in a two-year college when they meet the admission threshold. Combined, these differences with Hoekstra (2009) mean that the estimated return to enrolling in the respective colleges could be very different from one another. Not only do the students differ, but the counterfactual options are likely very different as well. Although not directly observed, Hoekstra (2009) likely estimates the effect of four-year college quality, while Zimmerman (2014) estimates the effect of public four-year versus two-year enrollment.⁴³ Zimmerman finds that marginal admits earn 22 to 27 percent more between 8 and 14 years after high school completion, especially for men and those on free and reduced-price lunch. He also finds that admission does not meaningfully change degree completion but it does improve time-to-bachelor's degree.

The analyses in Zimmerman (2014) closely aligns with analyses in Georgia (Goodman, Hurwitz, and Smith, 2017; Smith, Goodman, and Hurwitz, 2020). The primary distinction is that they study the SAT admission cutoff for the entire four-year public university system, not just a single college. Most students in Georgia attend a public institution, and the two-year, four-year margin also is a common consideration, making the results of the study relevant to a large portion of students.

⁴³ Zimmerman (2014) also estimates an instrumental variables equation where the endogenous variable is enrolling in any public four-year college in Florida, which more closely resembles the tradeoff between two-year and four-year public colleges. But there are few changes in the estimates to enrolling in FIU because few students forgo other in-state public colleges for FIU.

The authors use SAT administrative data linked to NSC for college enrollment and completion. They merge these data to TransUnion credit bureau data that allows them to measure predicted household income along with credit scores and loan balances and repayment.⁴⁴ The SAT minimum at the university system is 400 for SAT math and 430 for SAT verbal, which is about the 25th percentile of the score distribution. Economically disadvantaged students just barely eligible for admission that enroll in a Georgia public university increase their probability of earning a BA in six years by 41 percentage points. Furthermore, predicted household income at age 30 increases by 20 percent overall and by 40 percent for students from low-income high schools. Though imprecise, they find few impacts on credit scores, loan balances, or repayment. Similar to Zimmerman (2014), the counterfactual largely consists of enrolling in a two-year college, but a small fraction of students missing the cutoff go to private four-year colleges, out of state schools, or do not enroll in any college. These complications highlight the challenge of estimating the returns to two-year versus four-year colleges in the U.S.: students have a large choice set, so the counterfactual to the observed outcomes is rarely a single college or even a single college type.

Kozakowski (2020) is similar to the Georgia context in considering the returns to a state's entire public four-year system: Massachusetts. Using the minimum high school GPA admission criteria, she finds that low-income students with a GPA above the minimum are 58 percentage points more likely to enroll in a public four-year college in MA than those low-income students with GPAs below the threshold. Many of these students would have enrolled in two-year college, but a near equal amount would not enroll in any college. A slightly smaller fraction of those compliers would have enrolled in a private four-year college, to which there are over five times as many as there are public colleges in the state. Compared to this complicated set of counterfactual enrollments, she finds that these marginal enrollees in the four-year public colleges are 15 percentage points more likely to obtain a BA degree and earn 26 percent more 8-14 years after initial enrollment.

The set of RD papers all find economically meaningful impacts on the return to college quality, albeit in different contexts. These studies estimate the LATE of the return to different types of college types and quality. The complier populations may differ from the overall population. In particular, the compliers in these studies are the least academically qualified

⁴⁴ This is the only paper using RD methods that does not rely on state UI data for earnings and demonstrates a potential path for other researchers.

students at these universities. While this can make RD studies difficult to compare to other studies and reduce their generalizability, the LATE has important policy implications. In the RD papers in Florida, Georgia, and Massachusetts, many students were trading off public two-year and four-year colleges. This common margin is one that can be influenced by policy. For example, states may expand capacity at public four-year colleges, and doing so could draw in students from two-year colleges (or vice versa). The expected effect of such a policy on students is identified by the LATE in RD studies. Similarly, free community colleges and promise programs are likely to incentivize students who otherwise would have attended public four-year colleges to enroll in public two-year colleges (the diversion margin), and the RD LATE is applicable to understanding the returns these students could experience.

4.2.5. Other Instruments and Policies

We next discuss two additional sources of variation that researchers exploit and their corresponding results: distance between a student's home and college and percent plans. Although the former relies on cross-sectional variation across geography and the latter relies on variation induced by a policy (or changes to the policy), they both connect to the returns to college quality literature through the assumptions underlying equation (5). That is, conditional on the instrument, the distribution of ability and preferences are the same on average among students enrolled in the college type under consideration and students enrolled in the counterfactual colleges. We discuss each in turn, including various diagnostic tests to satisfy the identification assumptions and whether the assumptions are likely to be satisfied.

Card (1995) explores distance between students' home and nearby colleges as an instrument for college enrollment. Several authors have since adopted this approach as an instrument for college type and quality, often using variants such as distance to two-year and/or four-year colleges (e.g., Long and Kurlaender, 2009) or average quality of colleges within a certain radius (Long, 2008). The identifying assumption embedded in this approach is that the relative distance only has an impact on college completion and earnings through college enrollment or choice and not a direct effect on completion or earnings. Card (1995) discusses three potential violations of this assumption, even after controlling for standard background variables.⁴⁵ First, families that place a strong emphasis on schooling may choose to live near colleges and these

⁴⁵ More thorough critiques of the approach are discussed in Bound, Jaeger, and Baker (1995) and Staiger and Stock (1997).

families may have some unobserved ability or motivation that is correlated with earnings. Second, nearby colleges may impact the quality of primary and secondary schools. Third, there may be unobserved wage premiums and labor market conditions surrounding the colleges. Data advances have allowed researchers to make headway on the empirical relevance of some of these violations.

Returning to Mountjoy (Forthcoming) but focusing on the two-year versus four-year (i.e., diversion) margin, the author controls for distance to two-year and four-year colleges. Furthermore, he controls for neighborhood levels of urbanization and commuting zone fixed effects to mitigate the second and third concerns with the exclusion restriction listed above. However, the first concern is more difficult to control for so he implements a series of empirical tests and finds that the distance instruments are conditionally uncorrelated with pre-collegiate student test scores. Whether there are additional and unobserved abilities, motivation, or preferences that dictate where students and families choose to live is an open question and a drawback to the method. But datasets with an extensive set of variables are increasingly available and have the potential to further mitigate concerns with the identifying assumptions.

Mountjoy (Forthcoming) finds that the students on the diversion margin who are induced into two-year colleges instead of the counterfactual four-year college are less likely to complete a degree, and he finds suggestive evidence of lower earnings. Specifically, diverted students are 18 percentage points less likely to earn a BA degree, which corresponds to a negative but imprecisely estimated \$500 decrease in quarterly earnings around age 30, which is at least a 5 percent decrease relative to the mean quarterly earnings in the sample. These results are smaller in magnitude than the regression discontinuity estimates but along with being a different state, the complier populations are likely very different from one another.

Next, we discuss the use of percent rules that govern admissions in a state's public universities. Using the Texas example, the top 10 percent of a high school graduating cohort receive automatic admission to all in-state public colleges. The top 10 percent of a high school's cohort is determined by high school GPA. These policies often are implemented as an alternative to race-based affirmative action, with the intention of increasing access to high quality colleges for students from high schools that historically send few, if any, students to those colleges (e.g., UT-Austin and Texas A&M). Since the students from underrepresented high schools tend to have lower standardized tests scores and fewer financial resources, they may struggle to gain admission to the most selective public colleges in the state. But within a high school, there are mechanically

10 percent of students who gain access to these colleges with a top 10 percent plan. While these policies have been shown to attract applications and the enrollment of minority and economically disadvantaged students, there are still large disparities along these dimensions in college enrollment and success (Black, Cortes, and Lincove, 2015; 2020).

Percent plans provide an opportunity to estimate the returns to college type and quality through a regression discontinuity design. Intuitively, researchers can compare students just above and just below the percent threshold with the standard assumption that there is no manipulation. Uniquely, there are few concerns about density around the percent threshold because the assignment mechanism is based on an ordinal ranking. Researchers also benefit from linking to NSC, since these policies can influence whether students stay in state and impact the probability of appearing in the state administrative dataset for college or employment. However, it still is important to test for balance across the percent threshold because students who just barely get into the top 10 percent of a high school cohort may be systematically more sophisticated, motivated, or even aware of the policy, which may be independently correlated with college enrollment, completion, and earnings. This would be a violation of the identifying assumptions required to estimate a causal effect as expressed in equation (5).

Daugherty, McFarlin, and Martorell (2014) study the Texas Top 10 Percent Plan adopted in 1997. They link data from a single large school district in Texas to NSC for college enrollment and completion outcomes. The school district predominantly serves Black and Hispanic students who are underrepresented at the top public colleges in the state. That partially explains why their RD estimate shows that being in the top 10 percent of their high school class increases the probability of enrolling in one of the two flagship universities by 60 percent. Unlike the RDs with minimum admission criteria, they do not find that students are changing the quality of institution in which they enroll. Most students are drawn into the flagship from private or out-of-state four-year colleges. Although the study focuses on enrollment, it offers a path forward for similar RDs that track completion and earnings.

Black, Denning, and Rothstein (Forthcoming) also explore the impacts of the Texas Top 10 Percent Plan. They use state administrative data, which limits their ability to perfectly assess class percent because high school GPA are not included in these data. Hence, they are unable to estimate an RD. Instead, they exploit the introduction of the policy in 1997 through a difference-in-difference (DID) framework. They find a set of students who are high-achieving but in high

schools that do not send many students to the flagships, such that the introduction of the plan is likely to grant them access and induce enrollment (“pulled-in”). Conversely, they find a set of students who have high absolute achievement but who do not score relatively highly within their more affluent and advantaged high schools (“pushed-out”). The DID compares the pulled-in (and pushed-out) students before and after the policy to students who are neither pulled-in nor pushed-out. The authors find suggestive evidence of increased earnings 7-9 years after enrollment for the pulled-in students. They find little evidence of an earnings penalty for more advantaged students who are pushed-out of the flagship. These modest earnings effects are consistent with Daugherty, McFarlin, and Martorell (2014) in that the change in college quality may be modest for students considering the flagships in Texas. However, Black, Denning, and Rothstein (Forthcoming) is unique from all the previous four-year college research discussed in that it exploits a policy change. The authors construct control groups that have similar characters to the treated groups through predictive models. The biggest threat to whether the control groups have the same distribution of abilities and preferences to the treatment groups is not cross-sectional selection issues but whether different cohorts’ abilities and preferences differ from one another even in the absence of the policy change. Narrower windows of time may help alleviate such a concern but at the expense of sample size, precision, and potentially unwanted cross-cohort spillovers.

Lastly, between 2001 and 2011, California implemented a percent program with an important difference from the Texas policy. Students in the top 4 percent of their high school class only receive an admissions boost, not automatic admissions, to the University of California (UC) *system*. The UCs are the most selective public colleges in the state. Bleemer (2021) uses this policy to estimate the return to enrolling in more selective colleges. He links UC applicants between 2001 and 2013 to NSC and state earnings data to estimate an RD around the 4 percent threshold, finding that barely eligible students from “less-competitive” high schools are 12 percent more likely to enroll in one of the top four UC colleges. These are among the most selective colleges in the nation. Students forgo enrolling in less-selective four-year colleges but also two-year colleges – a consistent finding in this research. He then shows that these students induced into the top four public colleges are 30 percentage points more likely to earn a BA in five years and increase annual wages by as much as \$25,000. The magnitude of these results are noticeably large and consistent with the RD studies discussed above that induce students to meaningfully shift the quality of the college in which they enroll.

4.2.6. Student-College Match

The evidence discussed above generally shows a positive return to college quality, however the treatment effects vary across different types of students. Andrews, Li, and Lovenheim (2016) is one of the few papers to estimate distributional effects of the return to college quality. More commonly, researchers consider subgroup analyses or interaction effects. For example, Ge, Isaac, and Miller (Forthcoming) find a stronger return to earnings from attending a college with a higher average SAT for women than men. These subgroup analyses are informative (and encouraged), but they tend to be limited in theoretical underpinnings and statistical precision. One exception is the literature on student-college academic match.

Research on student-college match generates a metric of the alignment between the student's academic ability and the average academic ability of enrollees at the college to which the student applies or enrolls. This metric thus focuses on academic match and not other dimensions of congruence, such as financial or social fit, which likely play a role in longer-term outcomes as well. The specific way in which researchers measure match varies across papers. Some studies use high school GPA and course taking, along with standardized test scores, to predict the probability of being admitted to a college of a certain quality (e.g., Bowen, Chingos, and McPherson, 2009; Smith, Pender, and Howell, 2013). This approach provides a measure of student access to different selectivity levels that can be compared to the observed selectivity level. The result is a categorical measures of match – undermatch and overmatch – where undermatch refers to students with higher academic ability than the college to which they enroll and overmatch is the opposite. Other papers generate a continuous measure of match by comparing a students' percentile of academic ability to the student-weighted percentile of college quality to which they enroll (e.g., Dillon and Smith, 2017; 2020).

Studies find substantial amounts of mismatch regardless of the specific measure employed. Using two nationally representative samples, Smith, Pender, and Howell (2013) find that about 40 percent of students enroll in an institution that is of lower quality than one to which they could be admitted and attend. A portion of this undermatch is driven by relatively high-ability students enrolling in two-year colleges or no college at all, and it is driven by students from lower-SES backgrounds. Dillon and Smith (2017) use data from the NLSY97 and find substantial amount of both undermatch and overmatch. They show that both dimensions of mismatch are driven by student application and enrollment decisions, rather than by college admission decisions. Financial

constraints, information, and the public college options students face also affect the types of schools into which students sort.⁴⁶

With an estimated measure of student-college academic match, most studies in this literature then examine how college match influences the returns to college quality. This is implicitly asking whether the returns to college quality are uniform across the ability distribution within a college or college type. There is reason to believe that the student-college match impacts the returns to college quality, over and above the effect of college quality itself. On the overmatch side, researchers argue that the level of coursework rigor targeted toward relatively higher academic ability peers is too daunting for students with lower incoming academic preparation. The lack of preparation may make it difficult to succeed in coursework, impact completion rates, or direct students to “easier” majors (Arcidiacono and Lovenheim, 2016). This is the same argument made by critics of affirmative action.⁴⁷ On the undermatch side, the relatively lower level of coursework rigor and low level of persistence among academically marginal students may lead to lower levels of human capital accumulation. Additionally, students at less-selective colleges are more likely to come from economically disadvantaged backgrounds and be first generation college-goers. Students with these background characteristics have lower completion rates and likely a different social and labor market network than relatively advantaged peers at more selective colleges. On the other hand, undermatched students have the potential to excel in coursework and graduate near the top of the class. The effect of mismatch is largely an empirical question to which we now turn.

Measures of match effects differ across studies, but nearly all of the research on this question relies on a selection on observables identification strategy. Light and Strayer (2000) is one of the early papers on student-college academic match in the U.S.⁴⁸ The authors use the NLSY79 to estimate a two-period model of college attendance and completion, where the first period models the decision to enroll in any college and the second period models the decision of which school to attend. Similar to previous studies, they conclude that college quality and student academic ability impacts college completion. They also conclude that the college match matters

⁴⁶ Hoxby and Avery (2013) show additional evidence of undermatch among high achieving, low-income students. See Section 6 for a further discussion of this paper.

⁴⁷ We do not focus on the affirmative action literature, which has been reviewed by several different scholars (Holzer and Neumark, 2000; Fryer and Loury, 2005; Arcidiacono and Lovenheim, 2016).

⁴⁸ Student-college match is considered in other countries as well, for example, in England (Campbell et al., Forthcoming).

for most students, however more so for low-ability students at the higher quality institutions (i.e., overmatch) than for high ability students at lower quality institutions (i.e., undermatch).

Dillon and Smith (2020) use both NLSY79 and NLSY97, making use of different match measures and sample restrictions. They employ a richer set of controls than do Light and Strayer (2000) and similar (prior) studies, which they argue provides a strong case for unbiased estimates in the selection on observables case. In support of this argument, they show that the coefficient estimates stabilize when adding successively more controls. This increasingly-common strategy in selection on observables research helps rule out some potential sources of bias, likely related to ability, but whether it removes selection bias as it relates to unobserved preferences that are correlated with earnings in particular is less clear.

Relative to the results in Light and Strayer (2000), Dillon and Smith (2020) find that ability increases degree attainment across all college quality quartiles and college quality monotonically increases degree completion across all ability quartiles. The most salient difference between the studies is that Dillon and Smith use long-term earnings as one of their outcomes. They find strong evidence of a college quality effect for all student ability levels. For example, each 10 percentile point increase in the quality of the first college enrolled is associated with an additional \$1,480 of annual earnings for students with the median ability in NLSY97. The match effects are smaller in comparison, are imprecise, and depend on the sample and outcome. The strongest evidence of the match effect appears in the NLSY79 cohort 20-30 years after initial enrollment.

Bowen, Chingos, and McPherson (2009) and Smith, Pender, and Howell (2013) focus on undermatch, the former in a single state and the latter with a nationally representative sample of high school students (ELS:2002). Each paper predicts the college selectivity to which students have access and compares this prediction to the observed enrollment. These papers also use a selection on observables approach but do not go particularly far in either claiming causality or assessing any bias. They include typical academic and demographic control variables and find that undermatch is strongly associated with low completion rates.

It is important to put the match estimates into context with college quality estimates. It may be the case that mismatch is detrimental to long-term outcomes. However, college quality is shown to be an important determinant of long-term outcomes too, and so a potential undermatched or overmatched student (or policymaker) should consider the net effects of college quality and match. The selection on observable approaches in the match literature typically estimate the two

effects separately – finding an unambiguous effect of college quality and a smaller and less precise estimate of the match effect (e.g., Dillon and Smith, 2020). On the other hand, some of the most convincing research on the returns to college quality also estimates net effects. For example, the RDs with minimum admission criteria estimate the LATE for students who were barely admitted to the four-year public colleges (Zimmerman, 2014; Goodman, Hurwitz, and Smith, 2017; Kozakowski, 2020). These students are by definition the lowest academically achieving students (i.e., overmatched), but each paper finds large positive effects on college completion and earnings. This means that the positive impact of college quality dwarfs any negative effect from mismatch for these students.

4.3. For-Profits

The number of students enrolled in the for-profit postsecondary sector increased from about 450,000 to 1.3 million between 2000 and 2015. The corresponding number of for-profit institutions increased from 789 to 1,262 – a 60 percent increase.⁴⁹ The increased prevalence of these institutions, combined with the poor educational and labor market outcomes among students who attend these types of schools, has drawn the attention of policymakers and researchers.

Returning to Table 1, we see that for-profits disproportionately serve disadvantaged and vulnerable populations. Fifty-two percent of for-profit students are URM compared to 36.9 percent and 30.5 percent at public and private not-for-profit institutions, respectively. Almost 70 percent of for-profit students receive a Pell Grant, an indicator of financial need, compared to 50.2 percent and 45.1 percent at public and private not-for-profit institutions, respectively. They also enroll relatively older students and more veterans than comparable institutions. For example, for-profit students are on average 24 years old, compared to 22 years old in two-year colleges (Cellini and Darolia, 2017). For-profit students additionally have lower completion rates,⁵⁰ longer time-to-degree, lower earnings, and much higher student loan default rates than students at similarly-selective non-profit colleges. For example, the 2018 three-year cohort default rate for federal

⁴⁹ Source: NCES' 2017 *Digest of Education Statistics*.

⁵⁰ As shown in Table 1, two-year for-profit completion rates are much higher than are two-year public completion rates (60.3 vs. 30.4 percent). Hence, not all outcomes in for-profits are worse than among the non-profits. This highlights the importance of understanding whether the return to these degrees differs across sectors.

student loan borrowers who attend a for-profit institution was 11.2 percent in 2018, compared to 7 percent and 5.2 percent at public and private not-for-profit institutions, respectively.⁵¹

Additionally, for-profit colleges are more vocationally-oriented than public two-year or four-year colleges. This often corresponds to a certificate rather than a degree, but even the degree programs tend to be vocation-oriented (e.g., business administration, IT, nursing). As we discuss in Section 5.3, field of study can contribute as much to earnings as college quality. Whether the field of study is part of the treatment effect of enrolling in a for-profit is not immediately clear, but this has important implications for the most appropriate methods used to estimate the effect of for-profit enrollment. For example, if a student chooses a for-profit and as a consequence, discovers a passion for IT that would never have been discovered by attending a not-for-profit college, comparing students in IT programs across sectors can lead to biased estimates on the return to enrolling in a for-profit. Alternatively, if a student has a pre-existing passion for IT and decides between a for-profit and not-for-profit, then comparing students in IT across sectors is more appropriate. This type of information often is not available to researchers. If institution choice affects major choice, then it is more straightforward to estimate the return to enrolling in a given type of college (with major choice as one of the mechanisms) than to estimate the effect of majoring in a given field across institution types.

The combination of for-profit institutions serving an economically and socially disadvantaged population, along with veterans, and those students facing relatively poor long-term outcomes makes studying the for-profit sector important and challenging. Are the worse outcomes among this population due to for-profit enrollment or does it reflect the selection of who attends a for-profit? The primary difficulty, as usual, is disentangling the selection effect from the causal effect of the institution. Table 6, adapted from Cellini and Koedel (2017), demonstrates the various methods implemented by researchers to address these selection effects, along with the results of those studies.⁵² The review of the literature shows remarkably consistent results – for-profit colleges do not improve the labor market outcomes of students relative to alternative enrollment

⁵¹ Source: U.S. Department of Education FY2018 Official National Cohort Default Rates. <https://www2.ed.gov/offices/OSFAP/defaultmanagement/schooltyperates.pdf>. Accessed online 10/19/2021. The three-year cohort default rate is the proportion of the cohort entering repayment in a given year that defaults over the subsequent three years. People enter repayment when they leave college through graduation or dropout without enrolling in another postsecondary institution (e.g., graduate school).

⁵² Cellini (2021) reviews the literature with some recent developments but finds no substantive movement in the overall assessment of for-profit colleges.

options at the same level (i.e., two-year and four-year). The completion and earnings effects are often negative relative to some comparison group but at times they are zero (e.g., Lang and Weinstein, 2013). However, the much higher tuition in for-profit colleges makes the return on investment for for-profit colleges substantially lower than the alternative. There are a few studies that find positive returns, but that is usually in relation to no college enrollment.

The range of methods in Table 6 demonstrates that the search for exogenous variation in access to and enrollment in for-profits remains elusive. Similar to two-year colleges and some four-year colleges, the open access nature of these institutions precludes using admission criteria to generate exogenous variation in for-profit enrollment, and there are no apparent rationing issues in the for-profits. This may be by design, since it has been argued that for-profits act as a release valve to absorb overcrowding in other sectors (Deming, Goldin, and Katz, 2012). As such, identification strategies such as OLS, matching, and individual fixed effects are prominently displayed in Table 6. We briefly discuss the papers in the context of these methods, along with the few papers that use instruments or experiments.

Several papers use selection on observables methods: regression and matching (e.g., Deming, Goldin, and Katz, 2012; Denice, 2015). The estimation strategy assumes the choice of a for-profit over a not-for-profit is conditionally random, which is a relatively strong assumption. Along with standard potential issues of selection, for-profits spend much more on advertising than not-for-profits (Cellini and Chaudhary, 2020). This unobserved factor allows students to be influenced by advertising, which could be correlated with longer-term outcomes.

A few papers that find positive returns to for-profits use an individual fixed effects model (e.g., Cellini and Chaudhary, 2014; Jepsen, Mueser, and Jeon, 2016), which has some advantages relative to other contexts. The estimation strategy is often criticized for lacking generalizability to the broader population of students with no work histories (e.g., students directly from high school). But in the context of for-profits, many of their students are older, with work histories that make individual fixed effects an appealing approach to identify the effect of for-profit enrollment on earnings. These papers demonstrate how researchers make use of longitudinal data when evaluating for-profits but also highlight the relevance of the control group. For example, using a state's administrative data on for-profit enrollees, Jepsen, Mueser, and Jeon (2016) only examine people who are eventually treated, comparing their earnings before and after enrollment. This implicitly identifies the effect of for-profit enrollment relative to no college attendance. Using

NLSY97, Cellini and Chaudhary (2014) set up a difference-in-difference analysis where the control group consists of people who never enroll in college compared to for-profit enrollees (before and after enrollment). This means that the estimates of for-profit returns are relative to high school graduates with no college experience. Both of these papers find that for-profit colleges improve outcomes relative to non-college attendees. While certainly of interest, it also is important to understand how for-profit enrollment affects outcomes relative to enrolling in a non-profit (typically public) institution.

Cellini and Turner (2019) offer a leading example of the fixed effects strategy that assesses a variety of control groups for the treated for-profit enrollees. They have U.S. population level data from the IRS and the Department of Education. Their comprehensive data allow them to use individual fixed effects and exploit changes to earnings before and after enrollment in a for-profit institution. However, they go a step further and compare those changes to three different control groups: public college enrollees, a matched sample of public college enrollees, and a group of young workers who never attend college. One value of this approach is that the most appropriate control group is not clear, and so they provide a range of estimates. Another benefit of this approach is that we are interested in the effect of for-profit enrollment relative to both non-enrollment and to enrolling in the public sector. Indeed, it could be the case that for-profit colleges induce positive returns relative to no college enrollment, while the returns are lower than those experienced by public sector students. The authors find consistent results across samples. They show that “certificate-seeking students in for-profit institutions are 1.5 percentage points less likely to be employed and, conditional on employment, have 11 percent lower earnings after attendance than students in public institutions.” The results hold for most fields of study and for men and women. They also find that the outcomes are relatively worse among for-profit colleges that are mostly online or multi-campus chains. Their results further suggest that for-profit students do not experience earnings increases relative to a matched sample of non-college attendees, which suggests a large negative return net of tuition to for-profit enrollment.

Armona, Chakrabarti, and Lovenheim (Forthcoming) use an instrumental variables strategy to estimate the effect of for-profit enrollment relative to public enrollment on student debt and default as well as on labor market outcomes. Their instrument is comprised of a Bartik-style shift-share labor demand instrument interacted with the share of for-profit schools in a commuting zone. The thought experiment underlying their approach is that students are more likely to enroll

in college when labor demand is low, and when there are more for-profit colleges in a local area they are more likely to enroll in a for-profit. They show this is the case and that the instrument does not alter the distribution of students at for-profit versus public colleges. Examining outcomes separately for two- and four-year schools, they find that for-profits induce students to take out more loans, incur more student loan debt, and default on those loans at higher rates relative to similar public college students. They additionally show evidence that for-profits are more expensive and that they lead to worse labor market outcomes, which helps explain why loan origination and default both are higher among for-profit students.

To address concerns about student selection, two papers run resume audit studies to generate their own exogenous variation in for-profit credentials. Deming et al., (2016) and Darolia et al. (2015) send out fictitious resumes to employers and randomize whether workers attended a for-profit college on the resume. Darolia et al. (2015) find no difference in call back rates from employers between resumes that list for-profits and (relatively open access) public colleges, nor was there a difference with resumes that report no postsecondary education. Deming et al. (2016) focus on for-profit credentials from online colleges and find large reductions in call-back rates for resumes with a for-profit business BA and for health jobs.

Overall, the audit studies are consistent with the quasi-experimental studies and find little evidence of a positive return to enrolling in a for-profit relative to public college enrollment and even no college. There is scope for future work on the topic. Researchers should continue to look for additional sources of variation that induce students into for-profit colleges. There also is a complicated relationship between veterans and for-profits. Veterans receive federally funded educational benefits, such as the GI Bill and the Post 9/11 GI Bill, and the benefits are often spent at for-profit colleges.⁵³ This topic deserves further exploration.

4.4. International Evidence on College Quality

Much of the international evidence on returns to college quality is comparable to our discussion of four-year college quality in the U.S. in that students' options and typical counterfactuals frequently compare a relatively high quality college to a lower quality college offering similar degrees. One complication in assessing college quality in international settings is that applications often are major-specific, so some students' counterfactuals are colleges of

⁵³ See Barr et al. (2021) for a description of some of the military benefits for postsecondary education and how this frequently leads to veterans enrolling in for-profit colleges.

differing qualities and some are different fields of study within the same university. More generally, even if the application and enrollment to a high quality college is not tied to a field of study, part of the treatment effect is how enrollment impacts field of study.⁵⁴ This is carefully described and documented in Anelli (2020) and Kirkeboen, Leuven, and Mogstad (2016).

On the other hand, one of the benefits of studying college quality in international systems is that the admission process differs from that of the U.S. and often allows researchers to disentangle the selection of students into relatively high quality colleges from the causal impact of the college's quality. Some of those admission policies and resulting estimates are summarized in Table 7. International studies often exploit a centralized (to some extent) admission process, which is where we start our discussion.

It is again worth noting that our review of international college quality is far from complete. We focus on several papers that have compelling results and methods worth highlighting. We recognize that countries have varying types of postsecondary education, and these different levels, sectors, and types deserve more attention than what we can devote to them here. We focus on the quality dimension because of how much attention it has received in the literature, however quality often intersects with college type.

4.4.1. Admission Cutoffs in Centralized Systems

As described in Section 2, several countries around the world are centralized in that a government agency coordinates the admission and enrollment process. The specifics vary across settings, which dictates the identification strategies used and the relative contribution of each paper. Most of these centralized systems base admission on one (or two) metrics, which usually includes a nationally-administered exam score.

The combination of simple admission criteria (unlike many U.S. colleges) and a centralized sorting system has yielded some compelling results on the return to college quality. In Hastings, Neilson, and Zimmerman (2013), the authors study Chile's centralized system for 30 cohorts of students. Students rank in order of preference up to eight college and major combinations, which the authors refer to as a degree. The combination of entrance exam scores and GPAs determines whether they are eligible for admission to the degree, which sets up an RD for the colleges the students rank. Although the RD is similar in spirit to some of the U.S. based papers, there are

⁵⁴ This is also true in the U.S. but there is little evidence on how college choice impacts field of study, perhaps because most colleges in a students' four-year college choice set tend to have overlapping majors.

some noteworthy distinctions. First, there are over 1,000 RDs, one for each of the degrees, that can ultimately be stacked into a single regression. Second, the degree consists of the college and major, which the authors can separately assess because of the variation in quality within a major across the colleges. Third, the list of up to eight preferences provides an explicit counterfactual for each applicant. This last point is missing in almost all U.S. based studies but present in several international studies. As the authors show, the counterfactual used impacts the magnitude of the results.

Hastings, Neilson, and Zimmerman (2013) find that crossing the admission score threshold increases earnings by an amount equal to 4.5 percent of the average sample earnings. But degrees in the lowest quartile of selectivity only do so by 2 percent compared to 9.1 percent in the highest quartile. Once the counterfactual degree is considered, the estimates are 4.7 percent and 24.2 percent for the lowest and highest quartile of degree selectivity, respectively. Importantly, these are intent-to-treat estimates and do not directly assess the impact of enrollment in the degrees but rather the impact of crossing a threshold, which gives access to these degrees. The authors also discuss the importance of field of study, which we discuss in Section 5.3.

Kirkeboen, Leuven, and Mogstad (2016) conduct a similar study, focusing on the centralized system in Norway. The centralization allows the researchers to see a list of up to 15 ranked colleges and fields of study for each student. Admission is determined by high school GPA, again setting up an RD. Perhaps the biggest difference with the previous papers is that the authors develop an instrumental variables method to assess the impact of enrolling in colleges with different characteristics – the treatment on the treated effect.

The focus of Kirkeboen, Leuven, and Mogstad (2016) is on field of study, perhaps because, unlike in Hastings, Neilson, and Zimmerman (2013), the effect of field of study dwarfs that of institutional quality. The authors find that after controlling for field of study, there is a large payoff (about \$40,000 in annual earnings) to studying at The Norwegian School of Economics – a very selective university – relative to studying at The University of Oslo – the largest university. But all other institutions' effects are not (jointly) statistically different from The University of Oslo earnings. Overall, they find that roughly one-quarter of the variation in the payoff to institution-specific fields is explained by the institution, while the remainder is explained by field of study.

Moving beyond the mean impacts for students on the admission margin, some papers narrow their focus to specific colleges and outcomes. For example, Zimmerman (2019) revisits

the Chilean context, considering the impact of elite business-focused degree programs on a range of outcomes including leadership positions in publicly traded firms and earnings in the top 0.1 percent of incomes in the country. He finds that admission to these elite programs raises the number of leadership positions students hold by 44 percent and their probability of attaining a top 0.1 percent income by 51 percent. This is one of the few papers on the returns to college access and quality that moves beyond mean effects and considers distributional effects.⁵⁵ He also finds that the effects are entirely driven by males from high-tuition private high schools, showing that the treatment effects of college quality are not evenly distributed.

There is also a centralized system in China, whereby all college applicants take the *Gaokao* college entrance exam at the beginning of June every year. A few weeks later, with information on minimum admission cutoff scores, students rank up to five colleges (and three majors) in which they would like to enroll. They are granted admission to only one college. Again, there is an opportunity to understand students' preferences and counterfactual enrollment through this process, similar to the research in Chile and Norway. However, one challenge in this context is the availability of data. To overcome this hurdle, Jia and Li (2021) make use of a large-scale survey over six years. They oversample elite colleges and their students, which is a nationally defined set of 96 colleges representing about 5 percent of students. The authors also measure the minimum scores for the colleges, which varies by province and year, in order to estimate an RD on having access to an elite college. They find that those who score above the cutoff thresholds are 22 percentage points more likely to enroll in an elite college and earn 5.2-9.7 percent higher wages in their first job after college. Combined, this translates to a 28-45 percent increase in early wages attributed to enrolling in an elite college, which they show cannot be attributed to industry, occupation, or job location.

4.4.2. Admission Cutoffs in Decentralized Systems

There are several papers that employ RD methods in international settings without a centralized system. Unlike the centralized systems and more in the spirit of Hoesktra (2009) and

⁵⁵ Andrews, Li, and Lovenheim (2016) estimate the quantile treatment effects of college quality in Texas. Rodríguez, Urzúa, and Reyes, (2016) examine heterogeneous effects in Chile. Although the average impacts are positive in Chile, some people experience negative returns to college quality. No such negative effects are evident in Texas, however.

Zimmerman (2014) that focus on a single college, these studies provide small but well-identified and important pieces to the overall puzzle of assessing the impact of college quality on students.

Anelli (2020) exploits a minimum admission exam score at a prestigious elite private university in Italy. He finds a discontinuous jump in the probability of graduating on time of 26 percentage points and in subsequent income of 38 log points. Once scaled by the first-stage, the impact of enrollment in the elite university on income is 58 log points. Anelli also shows that part of the impact is due to changing field choices, but even after accounting for those changes the institutional quality effect is still 41 log points. This high return more than makes up for the relatively expensive tuition at this university.

Canaan and Mouganie (2018) do not explicitly use an admission threshold but rely on the fact that students in France must pass a high school exit exam before attending college. Those who marginally pass on their first attempt enroll in higher-quality colleges. The authors go to great lengths to show that nothing else of substance changes at the threshold, including years of education. By doing so, they are able to rule out channels other than college quality, which is a more general strategy that researchers can employ when there is the potential for a policy or threshold to change multiple factors at once. The existence of multiple factors does not imply that researchers cannot assess college quality but rather, additional checks need to occur at the outset. If other things do change at the threshold, then researchers cannot disentangle quality from other policy dimensions and are typically left evaluating the overall impacts of the policy, not that of enrollment or quality *per se*.⁵⁶ Canaan and Mouganie (2018) find that the relatively low-skilled students who marginally pass the high school exit exam and see an increase in college quality also experience a 12 percent increase in wages.

In Colombia, entrance exams scores determine entry into colleges. In one of the earliest applications of an RD on college quality, Saavedra (2009) evaluates access to a top-ranked college and finds that those applicants just above the scoring threshold are 16 percent more likely to be employed one year after college (about 25 percent once scaled by the first stage), perhaps due to higher academic ability peers (0.5 standard deviations) and higher expenditures per student (40 percent). One of the unique aspects of this setting and country is that students take a college exit exam, allowing the author to assess human capital accumulation that results from marginal

⁵⁶ This is also true of other policies, such as a student loan policy that changes access to college type and quality in Chile (Bucarey, Contreras, and Muñoz, 2020).

admissions. He finds that marginal admits score 0.2 standard deviations higher on the exit exam. This result is in slight contrast to Sekhri (2020), who studies the returns to enrollment in India's public colleges, which are more prestigious than private colleges. Although the author finds meaningful differences in wages, she finds no impact on exit exam scores and concludes the wage premium is all due to college reputation effects.

4.4.3. Quotas

One identification strategy that surfaces outside the U.S. are admission quotas for certain groups of people. Quotas are a way to ration enrollment in a plausibly exogenous way, however they are illegal in the U.S. In practice, quotas are often complicated policies, especially if done at the national level across many or all colleges. But the institutional details provide opportunities for researchers to both estimate the impacts of the quotas on enrollment and examine the returns to enrolling in more selective institutions.

In 2004, the University of Brasilia implemented a racial quota intended to increase the representation of darker skinned and Black students. This race-based affirmative action is similar to affirmative action policies in the U.S., but the explicit use of quotas is distinct. Francis and Tannuri-Pianto (2012) examine the impact of this policy and show a clear improvement in college quality for the beneficiaries of the program. Similar to Black, Denning, and Rothstein (Forthcoming) in the context of top percent rules, the overrepresented group impacted by the quotas may have been relegated to lower quality institutions, which is another margin to evaluate.

At the national level, South Korea has a College Enrollment Quota Policy, which limits the number of first-time enrollees public and private colleges can matriculate. There are also regional quotas. Kim (2021) makes use of these quotas in relation to changing cohort sizes, which provide exogenous variation in four-year college attendance. The academically-marginal students who enroll in four-year colleges receive a 60.3-75.5 percent increase in hourly wages relative to the alternative, which is comprised largely of two-year college enrollment but also of no college attendance. One additional feature of this study and the rich dataset used is that the author is able to assess several non-wage outcomes, such as health insurance, job satisfaction, and workplace benefits, almost all of which show signs of improvements with four-year college enrollment. Other outcomes that are less grounded in economic theory, such as marriage, smoking, and drinking are noisier, and he is unable to draw strong conclusions about them.

In China, each college has a province-specific quota. For example, in 2016, Peking University planned to admit 31 science major students from Sichuan and 111 from Beijing, and those quotas dictate the relative competition that students from each province face for particular colleges, even if the students are otherwise similar (Pu, 2020). We are unaware of any research that exploits these quotas in China, but we expect they will receive more attention in the near future because they are a potentially fruitful way to assess the impact of colleges and college quality on longer-term outcomes.

4.4.4. Other Identification Strategies

There are several other notable methodologies applied across different world contexts that we highlight either because of the potential opportunities to mimic in other countries or as a way to compare to U.S.-based studies.

Broecke (2012) uses U.K. data and an approach similar to Dale and Krueger (2002), with some important differences. Specifically, the author controls for the set of colleges to which students apply and to which they are conditionally accepted. Improving on Dale and Krueger (2002), Broecke also observes the preference order of colleges, which can help resolve omitted variable bias concerns related to student preferences. The idea behind his approach is to compare two students with the same set of conditional acceptances but who in the end enroll in different colleges. The reason behind those differences in acceptances between students may be correlated with longer-term outcomes, however, which is a potential limitation of this method. One of the virtues of this type of approach relative to some of the RD studies is that researchers can examine a more complete range of college quality. Broecke (2012) finds that a one standard deviation increase in college selectivity, as measured by scores of exam enrolled students, translates to a 7 percent increase in earnings.

Another approach, similar to previously discussed methods, is to exploit within family variation in college choice. Lindahl and Regnér (2005) study 19,000 siblings in Sweden, many of whom make different college choices. Although the authors do not focus on college quality, they note that there are substantial differences in earnings based on college choice (there are very few colleges in Sweden), with the older colleges inducing the larger premiums. The same challenges that exist in the U.S. persist – why do siblings make different choices – but college in Sweden is largely free, which may remove some financial considerations, and high school GPA is very important in admissions, which eliminates the estimation challenges associated with the holistic

processes in the U.S. The authors also show the importance of using within family variation over cross-sectional variation, which is consequential. These results highlight the fact that registry and administrative data in some countries can be critical to reducing bias in the returns to college quality estimates.

Almost all the papers we have discussed in this section employ methods that are relatively data demanding, often requiring discontinuities, application portfolios, or siblings. However, there are quite a few papers that are somewhat less data demanding. Some of these studies are summarized in Table 7 and rely on a selection on observables approach. They consistently find positive impacts of college and college quality on outcomes (e.g., Li et al., 2012; Rodríguez, Urzúa, and Reyes, 2016), with a few exceptions, such as technical colleges in Colombia and Chile (González-Velosa et al., 2015). Other papers, such as Chen (2019), generate its own data through a resume audit study. Chen finds that employers in China are 13 percent more likely to call back a job applicant who went to a “very selective” college in China over the “least selective” colleges. He also finds that Chinese employers call back job applicants from very selective colleges in the U.S. 9 percent more than the least selective colleges in the U.S.⁵⁷

MacLeod et al. (2017) take a unique approach among all of the papers reviewed – the return to college quality is inferred from the rejection of an economic model and not directly estimated. Using a few cohorts of college graduates in Colombia, they present two findings that suggest college quality (or as they write, “reputation”) impacts wages. First, they find that once colleges introduced an exit exam, employers could rely less on the college’s quality. This is confirmed in the data and suggests that part of the return to college quality is based on reputation. Second, they show that earnings grow with experience faster at higher quality colleges than at lower quality colleges. But employer learning-statistical discrimination models suggest the opposite should be true, unless of course, employers value the reputational effects and there is in fact a relationship between college quality and earnings growth. This is potentially a part of the return to college quality.

5. The Return to Different Types of Education

⁵⁷ Chen (2019) also shows a higher call back rate for job applicants who went to college in China over the U.S., even between the most selective U.S. colleges and least selective Chinese colleges.

The prior section demonstrated that there can be substantial differences in the returns to postsecondary investments depending on the type of institution in which one enrolls. Once students are enrolled in a given college, they face a number of educational paths that determines the educational content to which they are exposed. Postsecondary education encompasses a wide range of different types of courses, degree lengths, and pedagogical methods, some of which contribute to the returns to institution type. However, even within a given college, students vary in their majors, the length of their degree or certificate program, the extent to which they are taking a narrow, focused curriculum or a broad, liberal arts curriculum, how tightly their courses are linked to specific needs in the labor market, and exposure to remedial coursework. To the extent that exposure to these different educational programs affects the human capital one accumulates in college, these curricular decisions could have a large impact on the returns students experience to their postsecondary investment.

While it is sensible to make a conceptual distinction between *what* one learns versus *where* one learns, we note that in practice these two decisions are often linked. For example, vocational training in the U.S. is typically only available at two-year (or less-than-two-year) colleges, while in Europe vocational training is done by polytechnic colleges. Hence, the decision to obtain vocational training is tied closely with the decision to enroll in an institution that offers these programs. Institutions also vary in the types of majors they offer as well as in the quality of specific programs.

In this section, we discuss what is known in the literature about the educational and labor market returns to specific types of education, distinct and separable from the returns to the types of institution in which students enroll. As the prior discussion highlights, these are not truly separable decisions, and it is likely that differences in educational offerings across college types are a central mechanism that drives the differential returns to attending different institutions discussed in Section 4. Relaxing the assumption that curricular decisions and institution enrollment decisions are separable as well as understanding how curricular variation contributes to aggregate differences in returns to institution type are important directions for future research.

We focus on four dimensions of students' education investments that receive attention in the literature: 1) the return to vocational education, certificates, and diplomas, 2) the effects of remedial education, 3) the return to college major, and 4) the return to postgraduate/advanced

degrees. This is not an exhaustive list of types of educational investment but highlights some of the main areas of focus in the literature and dimensions that impact many students.

5.1. Vocational Education, Certificates, and Diplomas

Vocational, or career technical education (CTE) is a prevalent feature of education systems around the world. Such training is distinguished from more traditional educational programs by focusing on developing specific skills that are linked to particular industries or professions. Prominent examples include nursing, information technology (IT), court reporters, dental hygienists, and pharmacy assistants. The length of time needed for a given vocational certification varies considerably. In the U.S., there are two types of certificates and diplomas: short-term certificates that require less than one year of schooling and long-term certificates that require more than one but less than two years of study. Vocational associate degrees are also prevalent, and these degrees typically require two years of full-time study. In 2018-2019, almost one million short- and long-term certificates were awarded in the U.S., which represent about 25 percent of all undergraduate credentials awarded. This understates the prevalence of vocational credentials, as many vocational associate degrees also are awarded each year.

In the U.S., vocational training in higher education is conducted almost exclusively by public two-year colleges and by two-year and less-than-two-year for-profit colleges. Some of the for-profit institutions are “trade schools,” such as cosmetology schools. Community colleges and for-profits are broader institutions that offer more traditional academic programs along with vocational options. In Europe and other parts of the world, vocational training at the postsecondary level occurs through polytechnic universities. Students typically earn a four-year degree equivalent to a BA but focused on a vocational area. A comparison of the two systems raises the question of what the optimal length of training is for different vocational programs. Comparisons of returns to vocational degrees across systems can shed some light on this question.

Another distinction between vocational and postsecondary degree programs is that the former typically enroll older students with substantial prior work experience. Indeed, these programs are often seen as a way to retrain workers who experience adverse labor demand shocks due to business cycle variation or from structural shifts in the U.S. economy. This is important because such workers are of immense policy interest and because older workers have less time over which to reap the returns to any credentials they earn.

5.1.1. Estimating the Returns to CTE Degrees

In a human capital model (Becker, 1962), the returns to postsecondary vocational programs are determined by the value of the skills these programs provide to students in the labor market. Because vocational students tend to be older and attend colleges in their local area, they exhibit low geographic mobility and the relevant labor market for them usually is their local labor market. Hence, the demand for specific skills in different labor markets likely contributes to the differences in returns to vocational programs across space. As we discuss below, the empirical literature is characterized by heterogeneity in returns across different vocational programs but also across different states. Differences in the demand for vocational skills across areas is one plausible reason for the variation in findings. Currently, there is little work linking measures of local labor demand to the returns to vocational training, which we view as an important future direction for this area.

Vocational programs may have different returns over the life course as well. Hanushek et al. (2017) argue that the short run effects of vocational training on employment may be large because people are gaining skills that are needed in the current labor market. However, these skills degrade, especially with technological change, and the specific training associated with vocational degrees may lead workers to be less adaptable to changing skill demand over time. This would lead to worse employment outcomes later in life. They present evidence consistent with this hypothesis for secondary vocational training. There currently is no evidence on how returns to postsecondary vocational training vary over the life course.

A growing literature has arisen that examines the educational and labor market returns to different CTE programs. There are two main empirical challenges associated with these studies. The first is determining the correct comparison group. There are two potential comparison groups: 1) college non-attendees or students' pre-collegiate earnings and 2) those enrolled or completing a more traditional AA or BA program. Both of these comparisons are of economic interest. The first comparison answers the question of whether these degrees lead to returns over and above simply staying in the labor market. Essentially, this comparison shows whether the returns to the human capital accumulated is larger or smaller than the loss of experience from being out of the workforce. The second comparison provides insight into how the returns to vocational training compares to the other educational alternative – a traditional AA or BA degree. The majority of both vocational and traditional postsecondary degrees are awarded by open-access institutions, and so those enrolling in a vocational program could have enrolled in an open access AA or BA program as well. Ideally, we would have evidence on each of these parameters to fully inform our

understanding of the returns students experience to vocational postsecondary investments relative to their past outcomes and with respect to their other options.

The second empirical challenge is the difficulty of establishing causal estimates of these various parameters. Comparing students who enroll in a vocational program versus those who do not is problematic, as these individuals almost certainly differ in their background characteristics in ways that relate to labor market outcomes as well. Even a dataset with rich background controls will be identified from the strong assumption that any differences between enrolled and non-enrolled students are uncorrelated with potential outcomes.

The same concern exists in identifying the returns relative to enrolling in an AA or BA program. Differences across students in program enrollment are likely to be correlated with unobserved characteristics that themselves influence labor market outcomes. What is needed to overcome these selection problems is some source of exogenous variation in vocational program enrollment. Currently, little progress has been made on finding such sources of variation.

Because of these identification concerns, the majority of the research in this area has focused on identifying effects of vocational programs relative to an individual's prior earnings. This would be a challenging strategy for analyzing the returns to traditional higher education programs, where students usually do not have pre-enrollment earnings. Since vocational programs tend to enroll those with prior work experience, examining changes in earnings or employment surrounding program enrollment is a more credible strategy. Researchers have typically conducted such studies with individual fixed effects, which account for fixed unobserved differences across those who enroll in different vocational programs. Hence, this strategy permits a comparison of the differential returns to various vocational programs in terms of program length and focus, but it does not allow one to easily compare these returns to those experienced by students in traditional education programs.

We now turn to an examination of the literature on the return to vocational certificates and diplomas. We first discuss the extensive body of research using individual fixed effects, and then we discuss the much smaller set of studies that use alternative methods.

5.1.2. U.S. Estimates using Individual Fixed Effects

A growing set of papers employ individual fixed effects to identify the return to sub-baccalaureate CTE programs. As discussed in Section 3, the individual fixed effects strategy is not without its own set of assumptions. There are two main concerns with this approach. The first is

“Ashenfelter’s Dip” (Ashenfelter, 1978). This phenomenon refers to the fact that those experiencing an adverse shock are more likely to enroll in a given program. In this context, workers who experience a job separation or an earnings decline may be more likely to subsequently enroll in a vocational degree program. This would cause a positive bias in the estimates, as the pre-enrollment earnings are artificially low. Additionally, earnings in the first several years after enrollment will be depressed because students are attending classes rather than working full time.

Researchers typically examine event studies that allow them to determine if there are pre-enrollment declines in outcomes and that permit an examination of outcomes several years into the future. If these event studies show evidence of pre-enrollment earnings declines, researchers often exclude the several years surrounding enrollment from the analysis sample. However, this requires a longer panel.

The second identification issue relates to secular growth in earnings. Earnings tend to rise with worker age, which can affect within-worker comparisons over time surrounding program enrollment. This issue has received less attention in the literature, outside of Stevens, Kurlaender, and Grosz (2019), who include individual-specific linear time trends in their individual fixed effects models. While these controls are helpful in addressing concerns related to secular earnings growth, they do not account for bias stemming from secular earnings increases that occur post-enrollment. It is, however, not clear where such variation would come from, and, therefore, the model in Stevens, Kurlaender, and Grosz (2019) is currently the gold standard in this body of work.

Jepsen, Troske, and Coomes (2014) provide some of the first evidence on the returns to certificates, diplomas, and degrees using an individual fixed effects approach in Kentucky.⁵⁸ They use state administrative data on all students aged 20-60 who enter a public two-year college from 2002-2004. These data are linked to quarterly earnings data from the state unemployment insurance system. In Kentucky, certificates refer to programs that require one year or less, and diplomas require between one and two years of study. They find that among men, earning an associate degree increases quarterly earnings by \$1,484 (24.2 percent relative to the mean), diplomas increase earnings by \$1,265 (20.6 percent), and certificates increase earnings by \$297 (4.8 percent). Among women, the wage effects for associate degrees, diplomas, and certificates

⁵⁸ See Section 4 for a discussion of this paper’s findings on the return to AA degrees.

are \$2,363 (55.7 percent), \$1,914 (45.1 percent), and \$299 (7.0 percent), respectively. These results are generally consistent with subsequent studies in other U.S. states, as shown in Table 8.

Table 8 demonstrates that, in general, the returns to sub-baccalaureate vocational degrees are positive. There are some notable differences across states, however, and in some cases the shorter degrees are associated with larger returns than longer degrees. Interpreting comparisons across states is complicated by potential differences in the mix of fields of study. In Table 9, we show estimates from five different studies that employ individual fixed effects estimators and examine effects by level of sub-baccalaureate degree and by field. We present estimates of earnings returns based on state unemployment earnings data for five different areas: Health, Business, Nursing, Information Technology (IT), and Public Safety (PS).⁵⁹ We also break down credentials into three types: associate degrees, diplomas (programs between one and two years), and certificates (programs of at most one year). We separate estimates by gender for all but the Bettinger and Soliz (2016) results, which are pooled (but still shown in each panel).

Several patterns emerge from the results in Table 9. First, degrees and certificates in health and in particular nursing generate large returns. Women in Kentucky experience a 121 percent increase in earnings due to an AA in health (including nursing), and the nursing premium among women in Kentucky is 103 percent for an AA. The returns tend to be lower for diplomas and certificates in health/nursing, but outside of Washington State they are positive and substantial in magnitude. Generally, the returns to a vocational health degree are quite large across degree types and states.

Second, there is wide variability in the returns within credential types across states, particularly outside of health/nursing. For example, a diploma in business has a return of 17.6 percent in Ohio but -21.7 percent in Kentucky among men. A certificate in IT raises earnings by 25.3 percent and a certificate in public safety increases earnings by 65.3 percent among women in Ohio, while in California and Washington the returns are near zero. Given the similarities in the data being used and the estimation method, this variability is surprising. No work has been done to try to explain this variability.

One explanation is linked to variation in demand for these skills across states. The demand for health workers and nurses tends to be high nationally, while the demand for IT or protective

⁵⁹ The Health estimates in Jepsen, Troske, and Coomes (2014), Stevens, Kurlaender, and Grosz (2019), and Bettinger and Soliz (2016) include nursing.

services workers varies more regionally. The implications of this explanation are important. If the variation in returns is due to variation in labor demand, then it suggests pairing vocational programs with local employers to match the skills produced with the skills demanded is of high value. It also provides a guide to students to help them determine which are the most high-return degrees in their labor market. Another explanation is variation in labor supply across areas: the supply of workers trained for specific occupations likely varies across labor markets. This explanation also has important policy implications, as it suggests it would be valuable for colleges to adapt their degree/credential offerings to avoid over-producing certain types of skills.

Third, less time-intensive degrees tend to exhibit lower gross returns, but this is not a universal finding. Indeed, some short-run programs exhibit much higher returns than longer programs in the same area. Adjusting for the cost of attendance (tuition/fees and foregone wages), the net return to many shorter degree programs is quite large. For example, a one-year business certificate in California has a return of 17.9 percent among men and 16.6 percent among women, which is larger than the return to a business AA or diploma for either group. Protective services certificates also tend to have high returns. The explanation for why the gross returns to shorter programs are larger than the gross returns to longer programs is not immediately clear. It is likely these different programs are linked to different professions within each area of study, and it could be that there are substantial compensating differentials that increase the desirability of these lower-return professions. While there is ample evidence documenting these patterns, researchers have yet to put forward a unifying theory that can explain these results. This is an important area of research going forward.

5.1.3. International Evidence

In addition to the large literature in the U.S. examining the returns to vocational credentials, there is a smaller set of studies examining returns in other countries and across countries. Böckerman, Hämäläinen, and Uusitalo (2009) present some of the clearest evidence on the returns to vocational degrees in Europe. They examine a reform in Finland that transformed vocational colleges into polytechnic schools. These polytechnic schools typically offer four-year degrees and are more similar to traditional universities, except that the degrees they offer are more vocational and the faculty are not engaged in research. Essentially, the Finnish reform turned institutions that offered sub-baccalaureate vocational training into those that offer longer, baccalaureate-level

vocational training. The polytechnic schools in Finland are similar to those throughout Europe, so their estimates are applicable to European vocational training more broadly.

The authors employ an IV strategy that divides Finland into different “catchment” areas based on historic enrollment patterns of upper secondary students to vocational schools. They then use differences in when local schools were converted to polytechnics. This identification strategy is facilitated by the fact that most vocational enrollment is local, which is the case in Europe and in the U.S. The supply of schools in the local area thus defines the relevant options for vocational students. This approach is based on the assumption that the timing of conversions is not correlated with unobserved differences across student enrollment cohorts in terms of their earnings potential.

For those in business and administration, the polytechnic degree leads to a large increase in employment and earnings relative to a vocational school degree. In the other fields they examine (technology/transport and social/health care), they find modest employment effects and no impact on earnings. Interestingly, the large return to health degrees in the U.S. is not evident in Finland, which could be driven by differences in labor demand for health professionals across the two countries. They estimate a different parameter from the U.S. studies, showing how the returns to a longer, more intensive vocational degree compares to the returns to a less-intensive degree. Additional evidence from Finland using individual fixed effects models akin to those from U.S. studies that compare pre- versus post-degree earnings finds returns to polytechnic degrees to be positive and large for both men and women (Böckerman, Haapanen, and Jepsen, 2018). The magnitude of the earnings returns is similar to those found in U.S. studies.

Brunello and Rocco (2017) use micro data from the Programme for the International Assessment of Adult Competencies (PIAAC) to conduct a cross-country examination of how the returns to post-secondary and tertiary vocational education compare to the returns to more traditional, academically-focused degrees.⁶⁰ The PIAAC data are unique in that they allow researchers to observe direct measures of adult skills as well as labor market outcomes. However, cross-country regressions tend to embed strong identifying assumptions, as it is not possible to fully account for heterogeneity in education systems and labor markets across countries. Furthermore, the data are not longitudinal and thus do not allow individual fixed effects, and there

⁶⁰ They use data from 17 countries: Australia, Austria, Canada, Czech Republic, Estonia, Finland, France, Germany, Ireland, Japan, Korea, Netherlands, Norway, Poland, Spain, United Kingdom and United States.

is no source of exogenous cross-country variation. Instead, they employ selection on observables methods. While the assumptions needed to interpret these results as causal are strong, they provide novel evidence on the relationship between vocational training and measured skills.

The main findings indicate that those with vocational postsecondary or tertiary degrees have lower literacy, numeracy, and problem solving scores than their counterparts who obtain academic degrees at these levels. The estimates tend to be larger for tertiary students than for postsecondary students. That basic skills are lower among those who receive vocational degrees is not unexpected, as vocational training necessarily trades off more general skill for specific skills. The effect on earnings thus are important for understanding whether vocational students are indeed lower-skilled rather than having a more focused set of skills. Among postsecondary graduates, they find that women with vocational degrees earn about 4 percent less than women with academic degrees, while for men there are no earnings differences. The skills gaps are similar for men and women, though, which suggests the skills differences are not driving the earnings effects. Among tertiary degree holders, the earnings differences are more substantial, at about 20 percent. These estimates are much larger than the skill gaps they document. They additionally show that vocational graduates are similarly attached to the labor force as academic graduates, however the former have more experience. An implication of this pattern of results is that the return to experience is not large enough to compensate for the lower return to vocational versus general education.

5.2. Remedial Coursework

Many college students enter the postsecondary system with insufficient pre-collegiate training. The majority of colleges and universities in the U.S. are open access, meaning that any student with a high school degree can attend as long as they pay tuition. This lack of academic preparation for college is important, as pre-collegiate academic achievement is strongly predictive of collegiate attainment. Figure 2, adapted from Bound, Lovenheim, and Turner (2010), shows the likelihood of enrolling in college (Panel A) and completing college conditional on enrolling (Panel B) by quartile of the high school math exam students are administered in the surveys. These patterns are shown for two cohorts that graduated from high school in 1972 (NLS72) and 1992 (NELS:88). Panel A shows that college enrollment has increased markedly for those in the bottom two quartiles of the math score distribution. Panel B demonstrates that the likelihood of completing college is strongly related to high school math exam performance, and the gradient has increased

over time. In the NELS:88 cohort, only 11 percent of college attendees in the bottom quartile and 30 percent in the second quartile obtain a degree.

The high dropout rates of students with low academic preparation for college combined with rising rates of college enrollment among less-prepared students has led many colleges to implement remedial programs that are aimed at preparing students for college-level academic work. About half of all U.S. postsecondary students take some remedial coursework (Scott-Clayton and Rodriguez, 2015), with a total cost to students of about \$1.3 billion (Boatman and Long, 2018).

Theoretical predictions of how remedial education impacts student outcomes are ambiguous. Scott-Clayton and Rodriguez (2015) provide a theoretical framework that highlights this ambiguity. They discuss three mechanisms through which remedial coursework could affect students: 1) skill development that prepares students for college-level classes, 2) discouragement that may stigmatize them and sends a signal to students about low likelihood of success in college, and 3) diversion from coursework that advances their progress towards a degree. The diversion mechanism is driven by the fact that remedial coursework does not contribute to progress towards a degree, but students still have to pay tuition. Hence, it could slow down student progress by diverting resources and effort towards non-degree courses.

Furthermore, as highlighted by Bettinger and Long (2009), there could be peer effects from effectively tracking students by pre-collegiate achievement levels. Grouping less-prepared students together could lead to adverse peer effects that can impede attainment.⁶¹ Conversely, tracking students could allow colleges to target more resources to less-prepared students.

There is no uniform model of remedial education, and each institution offers these courses in a different way. Students typically are allocated to remedial courses based on placement exams that are taken when students first enroll in college. Placement test cutoffs for remedial courses based on these exams form the basis for regression discontinuity studies of how remediation affects student outcomes. There is variation across institutions and states in how the placement tests are used to assign students to remedial courses. In some states (e.g., Florida), there are unified standards for all public institutions across the state, while in others (e.g., Ohio) institutions set their

⁶¹ This also implies that there could be positive peer effects on more academically-advanced students who are not in remedial courses. There currently is no evidence on this question, however.

own standards. Colleges also vary in terms of the subjects in which they offer remedial coursework and the number of remedial courses they offer.

Typically, colleges offer remedial courses in “core” subjects, such as reading, writing, and mathematics. Programs vary in terms of whether they focus exclusively on academic skills or whether they additionally seek to boost non-cognitive skills. There also can be different levels of remediation into which students are placed. In some schools, remedial courses are taught by academic departments (termed “mainstreaming”), while in others remedial courses are taught centrally. Proponents of mainstreaming argue that it better aligns the remedial and academic curricula, however centralization can help ensure that students are learning a consistent set of skills and can facilitate the targeting of other resources to less-prepared students.

Identifying the causal effect of college remediation is difficult because students do not randomly take these courses. The courses are targeted to less academically prepared students, and thus cross-sectional comparisons of outcomes among remedial and non-remedial students are likely to be biased by unobserved differences across these groups. No dataset contains sufficient information to fully control for the ways in which remedial students differ from non-remedial students in terms of their potential outcomes. As a result, the majority of this literature uses regression discontinuity methods, exploiting placement exam cutoffs that determine whether students must enroll in remedial classes. Under the assumption that students cannot manipulate their exam score in the vicinity of the cutoff, these models identify the causal effect of remediation on subsequent student outcomes.

The local average treatment effect from the RD design estimates the effect of remedial coursework for those who are just below the threshold. It is important to emphasize that this LATE may not generalize to less-prepared students who are farther from the cutoff. Indeed, students who barely qualify may experience more diversion, while students who are much less prepared could benefit more from skill development (and may experience less stigma). How students with different levels of pre-collegiate preparation are affected by remediation also is important to understand because it provides insight into where to place the remediation cutoffs. If students close

to a current threshold are harmed but lower-performing students benefit from remedial courses, the best policy response likely is to lower the cutoff.⁶²

Martorell and McFarlin, Jr. (2010) provide one of the first and most comprehensive RD analyses of the effect of remediation on educational outcomes as well as the only evidence in the literature on labor market outcomes. The setting for their study is Texas, which requires two- and four-year students to enter remediation if they score below a “college readiness” threshold on the Texas Academic Skills (TASP) exam. It is common for students to score below this threshold, with 21 percent of four-year and 40 percent of two-year students failing to demonstrate college readiness on the TASP exam.

They use administrative data on public higher education students in Texas linked to quarterly earnings data from the state UI database to estimate RD models of being placed in remediation. In both two-year and four-year colleges, being placed in remediation reduces academic credits attempted in the first year and overall. There is a negative estimate of the likelihood of graduating within six years of about two percentage points, from a baseline of 32 percent in two-year and 48 percent in four-year colleges. However, the estimates are not statistically different from zero at even the 10 percent level. They additionally find little evidence that students experience lower earnings seven or more years after enrollment, though as is typical with earnings estimates they are imprecise. Taken together, the results from Martorell and McFarlin, Jr. (2010) indicate that placement into remedial courses leads students to accumulate fewer academic credits, but there is little impact on degree attainment or earnings.

Estimates from other regression discontinuity studies are broadly aligned with those in Martorell and McFarlin, Jr. (2010) for educational outcomes. However, no other paper studies labor market impacts. Calcagno and Long (2008) employ an RD approach in Florida and find that students substitute academic courses with remedial courses, which increases short-run persistence but not degree completion. This pattern is aligned with the diversion mechanism. Scott-Clayton and Rodriguez (2015) find similar results in a large, urban community college system.

Bettinger and Long (2009) examine the effect of remedial education in Ohio using a somewhat different identification strategy. In Ohio, each institution sets their own remedial

⁶² Concerns over proper targeting of remedial education and over whether these programs are harmful to students have caused three states – California, Connecticut, and Nevada – to fully eliminate remedial postsecondary education. No studies of which we are aware study these policy changes.

education assignment rule using some combination of ACT/SAT scores, high school transcripts, and assessment exams. This creates variation in remedial placement across otherwise identical students. Using the fact that students who attend open-access colleges typically attend the one closest to them, they instrument for remedial education placement using the assignment rule of the nearest college to each student combined with ACT and high school transcript information. The LATE identified by this instrument differs from those in RD studies, most notably in the fact that there is variation in pre-collegiate achievement among the marginal population.

Similar to the results from the RD analyses, they find remediation leads to high short-run persistence but lower total credit hours completed. However, they find large, positive effects on BA receipt, with larger effects for English than math remediation. An innovation of this study is the ability to examine effects across students with different pre-collegiate achievement. They interact a remediation indicator with linear ACT score to examine heterogeneous treatment effects. They show that the persistence and completion effects grow with ACT score, which suggests that relatively higher-achieving students benefit the most from remedial coursework. The skill boost from remedial classes for these relatively higher-achieving students may be more effective in closing skill gaps between them and non-remedial students. In contrast, remedial coursework may not be sufficient to allow lower-achieving students to catch up. There also could be important nonlinearities in how pre-collegiate achievement interacts with remedial placement that are not picked up by the linear interaction term. .

Boatman and Long (2018) extend our understanding of the heterogeneous impacts of remediation by exploiting a tiered remediation system in Tennessee. The Tennessee system requires those who are less academically prepared to enroll in more remedial courses, essentially generating multiple test score cutoffs for different levels of treatment. Comparing effects on students who only need to take one remedial course to effects among less-prepared students who take a more intensive remedial curriculum provides insight into which students are most affected by remediation. Their main findings indicate that students in the least-intensive remedial courses experience the largest negative effects of remediation on credit accumulation and eight-year degree attainment. The least-prepared students, who experience the most intensive remedial coursework, exhibit more modest negative impacts and in two-year colleges their educational outcomes improve. These results are consistent with a diversion effect among the relatively higher-achieving students, with a positive skill effect among less-prepared students.

The results in Boatman and Long (2018) are seemingly in contradiction to those in Bettinger and Long (2009), however the context differs in terms of the intensity of the remedial curriculum for less-prepared students. In Tennessee, lower-achieving students take more remedial courses, which could help them close the large pre-existing skill gaps they face upon college entry. Any skill increases experienced by the relatively more-prepared remedial students are insufficient to overcome the adverse effects of diversion. The results suggest that more intensive remedial education can help less-prepared students, but aligned with the studies discussed above there is evidence of an adverse diversion effect for those who barely qualify for remedial coursework.

5.3. College Major

One of the most important decisions college students make is their course of study, or major. College major choice determines some, but not all, courses students take, and this decision therefore influences the type of human capital students accumulate in college. Different majors also lead to different careers. For example, a student majoring in chemical engineering will gain different knowledge than a student majoring in early childhood education. It would be difficult for a chemical engineering major to teach 2nd grade after college, while the early childhood education student would be qualified for such a position. Conversely, the chemical engineer would be qualified to work for an engineering firm after college, while the early childhood education major would not. The return to different majors also can be realized within a given occupation. For example, the skills learned as an economics major may make economics majors more productive as management consultants than history majors. These examples highlight that some of the returns to different majors flow through occupational choices, while some of the returns are within occupations.

The return to field of study is part of the broader question of how the courses students take in college affect earnings. College major is a proxy for the courses students take, but there still is likely to be substantial variation across students in their coursework even conditional on major choice. The types of courses students take vary considerably across postsecondary institutions and across students within a given institution. Many colleges and universities have required course distributions that are motivated by perceived value to students of receiving a broad education. For example, at Cornell University, Arts and Sciences students must complete courses in arts, literature, and culture, biological sciences, ethics and the mind, global citizenship, historical analysis, physical science, social difference, social sciences, statistics and data science, symbolic

and mathematical reasoning, and foreign language in addition to their major to graduate. However, at Amherst College, there is no core curriculum, and students can take any courses they would like outside of their major requirements.

Do the returns to college major vary with the distribution of other courses taken? How important are these other courses conditional on college major, and how valuable is a core curriculum that ensures students receive some breadth in their collegiate training? The identification concerns associated with these open questions are difficult to overcome. Currently, there is no literature that examines these questions on the returns to course-taking behavior. With the increased prevalence of administrative data containing college transcripts, even descriptive evidence on earnings premia associated with course-taking behavior would be of high value. Because no research yet exists on the broader question of the returns to specific coursework, we focus on the returns to college major, on which there is a growing body of evidence.

Earnings differences across majors are large. The distribution of mean earnings by field of study among prime-age workers with a BA from the 2014-2019 ACS are presented in Figure 3. The majority of majors fall into the range of \$50,000 to \$100,000, but there is a long right tail with 22 of the 174 fields experiencing mean earnings of over \$100,000. Table 10 shows mean earnings for the 10 highest and lowest paying majors. The lowest-earning majors are early childhood education (\$39,542), cosmetology services and culinary arts (\$43,060), fine arts (\$44,744), and family and consumer sciences (\$44,892). The highest-paying majors are petroleum engineering (\$131,063), health and medical preparatory programs (\$128,713), actuarial science (\$112,827), and zoology (\$111,240).

Mean earnings differences across majors cannot be interpreted as the return to majoring in a given area because of student selection into different majors. Using the terminology from Section 3, students sort into majors based both on ability (A) and preferences (Q). For example, those who major in petroleum engineering are likely to have more mathematical and analytical skills than are those who major in early childhood education because of the different curricular demands of these majors. If these skills are independently valued in the labor market, some of the earnings differences across workers with different majors will reflect these different underlying skills rather than their field of study. Because of the strong link between majors and occupations, students in different fields are likely to have different occupational preferences as well. Petroleum engineering

students may have different preferences over professions than early childhood education students, and these preferences will lead to earnings differences even absent major choice.

Estimating the return to college majors has received a lot of attention among economists. Many studies attempt to overcome these selection problems using selection on observables methods (e.g., James et al., 1989; Grogger and Eide, 1995; Hamermesh and Donald, 2008; Kinsler and Pavan, 2015). These methods are unlikely to fully account for the different abilities and preferences that will themselves impact future earnings among students who select different majors. A small body of work has attempted to estimate the return to majors using structural models that explicitly model student selection combined with instruments for student major choice. Arcidiacono (2004) is a prominent example of such studies. He estimates a model that is flexible enough to allow students to experience major-specific returns to ability and is identified predominantly from period-specific shocks that allow students to react to information they are receiving about their own preferences and abilities. He also employs state average earnings by major as an instrument for major choice, invoking the assumption that state-average earnings only affect future earnings but not ability or preferences. Using data from NLS72, he finds business and natural science majors experience the highest returns. Finally, several papers have employed regression discontinuity approaches that exploit admission cutoffs to different majors (Hastings, Nielsen and Zimmerman, 2013; Kirkeboen, Leuven, and Mogstad, 2016; Andrews, Imberman and Lovenheim, 2017; Bleemer and Mehta, forthcoming).

The research on the return to college major has focused exclusively on the mean. There has been little work done to understand effects on the distribution of earnings or on the distribution of returns across workers. How variable are the returns to college major across workers and what factors drive this variation? These are important questions because if the variance across workers is large, then the mean return is associated with substantial *ex ante* risk. Relatedly, understanding how different majors affect the within-worker variance of earnings over time is important. Some majors could lead to more stable earnings profiles, which is a missing consideration in the returns to majors literature. We highlight these issues as important directions for future research that are feasible to examine with large administrative datasets. Our discussion focuses on the mean returns to college major, however, because that is what the existing research estimates.

The literature on the returns to college majors has been reviewed in depth by Altonji, Arcidiacono, and Maurel (2016) and Altonji, Blom and Meghir (2012).⁶³ In this section, we provide an update on this literature since these prior reviews by focusing on several recent developments: 1) the emergence of regression discontinuity studies on the return to college majors, 2) studies on the role of occupational choices in driving the returns to major choice, and 3) lifecycle returns and skill obsolescence.

5.3.1. Regression Discontinuity Studies

As discussed in Section 2, regression discontinuity methods are a powerful way to overcome selection on ability and preferences. As long as students cannot determine on which side of the assignment discontinuity to locate, observed and unobserved attributes will move smoothly through the cutoff. This balances the distribution of characteristics among students on either side of the discontinuity. The challenge in the context of college majors is on finding discontinuous assignment rules that allocate otherwise similar students to different college majors.

As discussed in Sections 2 and 4, several European and South American college admission systems are well-designed for regression discontinuity approaches. In these systems, students apply to a college and a major simultaneously. These programs admit students based on college entry exam scores and high school performance. This creates implicit or explicit cutoffs, depending on student demand for each program. The two most prominent papers to use this method are Kirkeboen, Leuven, and Mogstad (2016) in Norway and Hastings, Nielsen, and Zimmerman (2013) in Chile. We focus our discussion on the former paper, as it is the most prominent illustration of this method and focuses on the importance of majors.

Using administrative data in Norway, Kirkeboen, Leuven, and Mogstad (2016) estimate RD models of program admission to identify the returns to college majors.⁶⁴ This feature of their study underscores that the LATE they obtain is likely to differ substantively from OLS estimates that identify the treatment effect on the treated. The RD model shows how earnings are affected by selecting a given major *relative to a student's next most desired major*. In contrast, the treatment effect on the treated measure identifies average differences in earnings across students who select different majors that may have little overlap in student interest. For example, we can compare

⁶³ These reviews also discuss the large body of work examining the factors that determine how students select different majors. We do not discuss this research because this review is focused on the returns to different collegiate investments.

⁶⁴ See Section 4 for a discussion of this paper's approach and findings with respect to college quality.

earnings of early childhood education students to petroleum engineering students, but few petroleum engineers are likely to have preferences for studying early childhood education if they could not get into their preferred program. The RD model, in contrast, restricts comparisons to majors that are connected through students' preference rankings. This is arguably a more relevant comparison, because we are estimating the returns to a given major relative to the relevant other majors for this group.

The main identification assumption using this method is that students cannot sort on the admission thresholds. This would be violated if students knew what the GPA cutoffs were and aimed for these cutoffs because of their interest in a specific program. This is made more plausible in Norway because the cutoffs are unknown and vary over time based on variation in the applicant pool and student demand. Typically, GPA and admission scores are difficult to manipulate with sufficient precision to violate the assumptions needed for RD models, which is why this method has become more prevalent in the literature.

Kirkeboen, Leuven, and Mogstad (2016) find a wide range of returns to different field choices with respect to the next-best alternative, with medicine, law and business exhibiting the highest average returns. On average, the return to humanities is negative. One of the most striking findings in this paper is that the returns to most majors relative to the next-best choice are positive. They argue that this result reflects sorting on comparative advantage: students know their own preferences and skills and sort into majors in which they have a comparative advantage. This sorting will generate a positive return to most majors relative to students' next-best option.

As discussed in Section 4, Kirkeboen, Leuven, and Mogstad (2016) also examine the return to college quality. Their main models impose additive separability between institution and major, which follows the rest of the literature. In sub-analyses, they are one of the few papers to relax the additive separability between major and institution type. The importance of college major could be larger or smaller at less-selective institutions, which is a potentially important source of heterogeneity in the returns to college that has not been sufficiently explored.⁶⁵ For example, the relative return to a liberal arts degree could be quite different at an elite liberal arts college than at

⁶⁵ Arcidiacono, Aucejo, and Hotz (2016) show that Black and Hispanic students who attend more-selective universities in California are less likely to complete a STEM degree than their observationally-equivalent counterparts at less-selective institutions. This highlights the importance of understanding the tradeoff between major and institutional quality as it relates to subsequent earnings. See Arcidiacono and Lovenheim (2016) for a more complete discussion of the quality-fit tradeoff in higher education.

a non-selective four-year school. Relaxing this assumption is an important direction for future research, which should be increasingly feasible with the availability of large administrative datasets.

While admissions cutoffs in countries with centralized, test/GPA-based admission systems are straightforward to implement, this method is less applicable in the U.S. system. Students applying to American universities typically apply to the university rather than to a specific major, and admissions is less rule-based on the U.S. Two papers have implemented this approach in the U.S. to study specific majors: business (Andrews, Imberman, and Lovenheim, 2017) and economics (Bleemer and Mehta, forthcoming).

Andrews, Imberman, and Lovenheim (2017) exploit the fact that at many universities in Texas, access to business undergraduate majors is restricted based on high demand. Many institutions in Texas have college GPA cutoffs for declaring a business major that are either explicitly set to regulate demand or are implicitly set based on student applications. A central empirical challenge in their study is that they do not know which institutions and years have cutoffs. Using administrative data in Texas, they search empirically for cutoffs associated with the business major by estimating RD models for each college and entering cohort of students from 1998-2008. Estimating a separate RD model for GPAs between 2.5 and 3.65 (with 0.05 increments), they characterize an institution as having a GPA cutoff for business of the t-statistic on any cutoff is above 2.5 in a given year.

The authors examine students who first enroll in college without declaring a business major, as these are the students who are subject to these policies. Using the cutoffs unearthed through statistical techniques, they then estimate RD models that identify earnings effects of business majors relative to what students would have majored in if they had not been able to choose business. A strength of this approach is that they are able to show the counterfactual distribution of majors. At the cutoff, the likelihood of majoring in business increases by about 5 percentage points, with the majority of counterfactual majors coming from economics (13.5 percent), STEM (51.6 percent), and Social Science (17.7 percent). Overall, they find little effect on earnings, although the standard errors are large. They also examine whether selecting a business major affects postsecondary outcomes, which is a potential mechanism for any earnings effects. They do not find any evidence that collegiate outcomes vary across business and counterfactual majors.

Bleemer and Mehta (forthcoming) conduct an RD analysis using data from economics students at University of California – Santa Cruz (UCSC). UCSC implemented a cutoff in 2008 whereby students with a GPA below 2.8 in the two introductory economics courses were generally excluded from declaring an economics major. Using earnings data from the California UI system, they find that majoring in economics increases subsequent earnings by \$22,000, or 46 percent. They are able to examine how majoring in economics affects occupational choice and find that this major shifts student preferences for business and finance careers. They estimate that about half of the earnings returns can be explained by changes in the industries in which students subsequently work. This is an important finding in suggesting that the returns to this major reflect both occupational selection effects and productivity effects within each occupation.

5.3.2. The Link Between Occupations and Majors

The findings in Bleemer and Mehta (forthcoming) underscore that occupational selection is a potentially important mechanism driving the returns to college major choice. Understanding how major choice affects occupational choices and how the returns to college majors varies by occupation is an area of growing focus in the literature. Simply put, these papers seek to understand whether the returns to college major are ubiquitous, or whether they vary across different occupations. If there is such variation, does it matter how close the occupations are in terms of skill requirements to the major?

Kinsler and Pavan (2015) provide one of the first examinations of how the returns to college major are mediated by occupational choice. They use data from the Baccalaureate and Beyond (B&B) Study, which is a nationally-representative sample of college seniors sampled in 1993 and then again in 1994, 1997, and 2003. Critically, this survey asks respondents whether their job is related to their field of study from college. While a self-reported variable, these responses are informative about whether workers perceive the skills they learned in college to be related to their post-collegiate occupation. They show descriptively that workers who report their occupation being closely tied to their college major earn more: 30 percent in science, 3 percent in business, and 11 percent in their residual/other major category.

That workers in jobs tied to their college major earn more could be because of the human capital acquired in college but it also could be driven by sorting on correlated characteristics. To estimate the causal link between closeness of the occupation to the major, Kinsler and Pavan

(2015) estimate a dynamic structural Roy model, where individuals are endowed with two types of human capital: math and verbal. Students are unsure of their underlying human capital and thus select majors based on noisy measures of their aptitude in these different areas. When they graduate, they enter the labor market and learn their human capital in both areas. Wages are determined by human capital in each area, major, and whether the major is related to the (endogenously selected) occupation. They estimate this model and find that the model generates similar premia for working in a major-related field. While they only examine three majors, their results are important in highlighting the relevance post-collegiate occupational choices in driving the returns students experience to postsecondary investments.

Altonji, Kahn, and Speer (2016) take a task-based approach to understanding how occupational selection affects the return to different fields of study. Following the seminal work of Autor, Levy, and Murnane (2003), they use the Dictionary of Occupational Titles that allows them to measure the abstract, routine, and manual task content of different jobs. They then map each major onto the set of occupations into which students with that major sort to calculate major-specific task measures. Using earnings from the National Survey of College Graduates (1993, 2003) and the ACS (2009-2011), the paper first documents that the earnings disparity across college majors has grown substantially over time. With the data on task requirements attached to each major, they estimate task-specific prices and decompose the change in the earnings variance across majors into the part due to changes in task prices and the part due to other factors. Changing task prices explain about two-thirds of the increase, suggesting that the value of human capital in certain college majors has risen and is realized by students taking jobs in the professions that require these skills.

Leighton and Speer (2020) take a related approach to analyzing the role of occupations in driving the returns to college major. Using ACS data, they first calculate a Gini coefficient of earnings premia for each major across different occupations. They argue that this measure captures the transferability of skills across jobs. The more unequal the premium across occupations, the less transferable are the skills associated with a given major. Leighton and Speer (2020) use this approach to separate majors into specific and general majors and then examine how the returns vary across these different groups. Descriptively, they provide novel evidence on how general or specific each major is, finding that education and nursing are the most specific and psychology, music, and philosophy are the most general. While their approach does not allow them to control

cleanly for selection into different majors, their analysis presents novel information on how the average wage premium associated with each major varies by the specificity of the major. Their main findings indicate that specific majors are more lucrative, with premia 5-6 percent higher than the average major and 15-20 percent higher than the most general majors. These results align with those in Altonji, Kahn, and Speer (2016) and Kinsler and Pavan (2015) in showing that professions that link most closely with the valuable skills learned in college exhibit the highest returns. Taken together, these papers indicate that a large component of the return to majors operates through occupational selection.

5.3.3. Lifecycle Returns and Skill Obsolescence

The returns to college major papers discussed above typically examine earnings during one's early-mid career. This is predominantly driven by data limitations, as there are not many datasets that allow researchers to link college major decisions to lifecycle earnings. An open question remains as to how the return to college majors varies over the life course and how predictive early returns are for permanent earnings.

Deming and Noray (2020) present some of the first evidence on this question, with a specific focus on the return to STEM degrees. The motivation for their analysis is that many high-paying professions attached to STEM also experience repeated and accelerated technological change. A computer programmer who graduates from college in the 1990s will find her skills out of date rather quickly unless she invests in on-the-job (or other) training. This “race between education and technology” (Goldin and Katz 2010) is an important feature of labor markets. Deming and Noray (2020) are the first to examine how this type of technological change affects the lifecycle returns to college majors.

They use data on job openings from Burning Glass Technologies, which is a large repository of online job vacancy ads in the U.S. The Burning Glass data include information on unique skill requirements that come from the text of the job ads. Using these reported skill requirements, they measure both old skills that used to be prevalent in ads but no longer are as well as new skill requirements that become more prevalent over time. This is a direct measure of technological change that embeds changing skills that employers are seeking in the labor market. They document that skill turnover is high: 29 percent of vacancies required a new skill. These shifts are not evenly spread over occupations. Computer and mathematical jobs require new skills at a much higher rate than jobs in education, law, and community and social services. These

changing skill requirements have important implications for the lifecycle returns to these professions (and the majors that feed into them). Returns to computer science may be high early in the career, but they then can fade as one's skills become more obsolete.

Using data from the 2009-2017 ACS, Deming and Noray (2020) show that those with majors linked to rapidly-changing occupations experience high returns early in the career and then declines over time. Much of the flattening of the experience curve for these majors is driven by exit from the high-paying but rapidly-changing occupations. Over time, as these workers' human capital becomes less relevant, they switch to occupations with flatter wage growth but with less technological change. This phenomenon is particularly strong for STEM fields, which experience extensive technological change.

5.4. Post-graduate/Advanced Degrees

Graduate degrees are becoming more prevalent in the U.S., as shown in Figure 4. In 2019, 833,706 Masters (MA) degrees were awarded by U.S. institutions, along with 187,568 doctoral degrees. These awards are up substantially from their 2001 levels, at 473,502 and 119,585, respectively. Much of the increase in MA degrees has come from education, health, and business. At the doctoral level, law (including JDs) and medicine (including MDs) are the most prevalent areas, and health/medicine is responsible for much of the increase over time in these degrees.

Figure 5 presents trends in the proportion of the U.S. population with an advanced degree, overall (Panel A) and separately by gender in Panels B and C. As of 2020, 14.1 percent of workers aged 25 or over had some form of graduate degree, with women being slightly more likely to hold an advanced degree than men (14.5 percent vs. 13.8 percent). About 10.5 percent of those 25 and over had an MA degree, 1.5 percent had a professional degree, and 2.1 percent held a doctoral degree.

These degrees can be quite expensive, especially MA degrees that typically do not offer much institutional financial aid and as a consequence, students completing a post-baccalaureate degree tend to take on substantial debt. In 2015-2016, those with any graduate degree, on average, had \$82,800 in student loan debt, \$71,000 of which was solely from graduate school.⁶⁶ Those with an MD, a professional health degree other than an MD (such as a dental or chiropractic degree), and a law degree had the highest levels of graduate student debt, at \$223,110, \$193,800, and \$131,700, respectively. Despite the high levels of debt, student loan default rates are low for post-

⁶⁶ These tabulations come from the *Digest of Education Statistics*, Table 332.45.

graduate students. The disparity between student debt levels and default rates suggests that those who obtain these post-graduate degrees have better labor market outcomes.

Mean earnings by postsecondary degree level are consistent with high returns to these degrees. In 2019, average earnings for workers 25 or older with a BA was \$59,134, while average earnings for those with an MA was \$71,631, with a professional degree was \$110,816, and with a doctorate was \$101,147.⁶⁷ Average earnings of physicians are particularly high, at \$343,600 and is over \$500,000 per year for surgeons and specialists (Gottlieb et al., 2020). The earnings premium of those with a post-graduate degree relative to a BA has grown over time. Between 1980 and 2012, the college-only wage premium (relative to those with sub-baccalaureate training) grew by 0.272 log points, while the post-graduate degree premium grew by 0.410 log points (Lindley and Machin, 2016), implying a relative premium associated with graduate degrees of almost 15 percent.

The high average earnings of those with post-baccalaureate degrees combined with low student loan default rates despite high debt levels is suggestive of high returns to these degrees. However, students sort into graduate school based on preferences for the high-paying fields linked to these degrees and academic aptitude for graduate training. These factors are likely to independently affect earnings, as the model in Section 3 highlights. The rising proportion of workers obtaining these degrees and the high earnings they experience has led to a growing literature that seeks to identify the return to post-graduate degrees. In this section, we discuss identification issues surrounding estimating the return to graduate degrees as well as the burgeoning research on the return to some of the most prevalent degrees: medicine, law, and business.

5.4.1. Identification Concerns

Section 3 laid out a model based on Altonji and Zhong (2021) to help clarify the identification assumptions invoked by different approaches to identifying the returns to a postsecondary degree. The original formulation of this model surrounded graduate degrees, which we briefly discuss here. All variables are the same as defined in Section 3, and we now introduce G_{it} , which is the graduate degree individual i holds in period t . Analyses of the return to graduate degrees typically condition on the undergraduate degree (c) as well. This leads to the following

⁶⁷ Source: <https://www.census.gov/data/tables/time-series/demo/income-poverty/cps-pinc/pinc-03.2018.html>.

expressions for the conditional mean of earnings and the treatment effect on the treated (Altonji and Zhong, 2021):

$$\bar{w}_{cgt}|G_{gt} = \sum_j \int_{A,Q} p_{cgt}(j|A_t, Q_t) w_{cgjt}(A_t) dF_t(A_t, Q_t|c, G_{gt}) \quad (7)$$

$$\bar{w}_{cot}|G_{gt} = \sum_j \int_{A,Q} p_{cot}(j|A_t, Q_t) w_{c0jt}(A_t) dF_t(A_t, Q_t|c, G_{gt}) \quad (8)$$

$$TT_{cgt} = \bar{w}_{cgt}|G_{gt} - \bar{w}_{cot}|G_{gt} \quad (9)$$

The assumptions underlying the identification of the TT parameter are very similar to those discussed in Section 3, however, we now condition on the undergraduate major as well. This can help account for some cross-sectional heterogeneity among students, although there likely remains substantial selection into graduate school across students within major.

There are three main approaches that the literature has taken to identify the returns to graduate training: selection on observables, individual fixed effects, and college-graduate program fixed effects. We previously discussed the assumptions embedded in the first two approaches. We emphasize here that these approaches require strong assumptions. In particular, the assumptions needed for identification using individual fixed effects is likely stronger in this context than in examining undergraduate degrees. Recall that for the individual fixed effects estimator, the pre-degree earnings need to accurately measure counterfactual earnings if the person did not get the degree. In the graduate school context, there are two main sources of bias that can violate this assumption. The first is that the pre-degree earnings could be lower because students take jobs that help prepare them for graduate school. For example, students may work in low-paying internships or research positions, or they may opt to take jobs in law or medical offices that are low-paying but help signal an interest in these professions. This makes pre-graduate school earnings a poor measure of counterfactual earnings for these workers and is likely to bias upward estimates of the return to post-graduate degrees.

Second, workers just out of college may learn about their abilities and preferences, which can affect their decision to invest in graduate school. This would cause $dF_3 \neq dF_1$ in the three period model. As an example, a student could take a job as a consultant after college, intending to get an MBA but realize she does not like this occupation. Instead, she may decide to attend law

school, however her pre-graduate school earnings reflect her initial preferences (for an MBA) rather than her preferences at the time of graduate school entry. This bias could go in either direction depending on how preferences shift for specific workers.

Altonji and Zhong (2021) and Altonji and Zhu (2021) employ a new estimator to examine the return to graduate degrees: college-graduate program fixed effects (FE-cg). This approach includes fixed effects both for the graduate degree one obtains and the undergraduate degree. In both Altonji and Zhong (2021) and Altonji and Zhu (2021), they include interactions between undergraduate and graduate degrees, allowing the effect of graduate training to vary by undergraduate degree. They also control for interactions between age and undergraduate degree, which accounts for some secular growth in earnings by BA field. The advantage to this model is that it does not restrict the sample to those with pre-graduate school earnings, as everyone with a given undergraduate degree (c) contributes to identification of the BA estimates. This helps relax data constraints, but it could be problematic to include those without graduate degrees if BA-only earnings among those with a given BA is an inaccurate counterfactual for those who eventually obtain a graduate degree. Given concerns about how preferences evolve with work experience, it also could be that average earnings among those with BA degree c is a better counterfactual for those who obtain a graduate degree than are the pre-graduate school earnings used in the individual fixed effects estimator.

Altonji and Zhong (2021) employ individual fixed effects and FE-cg models to estimate the returns to a wide range of graduate degrees using national data from the National Survey of College Graduates and the National Survey of Recent College Graduates. Altonji and Zhu (2021) estimate similar models using large, administrative linked K-12, higher education, and UI-based earnings in Texas. These are the only papers that examine a broad set of graduate degrees. They use a variety of different models and estimate returns to a large set of degrees, making their results difficult to summarize. As well, there are a number of papers that have estimated the return to specific degrees, such as an MBA or medical degree. Table 11 presents an overview of the findings from this research for three degrees: business (MBA), medicine (MD), and law (JD). We discuss each of these literatures in turn.

5.4.2. The Returns to an MBA

Table 11, Panel A presents estimates of the returns to an MBA. While all estimates are positive, there is a wide range, from under 5 percent (Arcidiacono, Cooley, and Hussey, 2008) to

0.210 log points (Altonji and Zhu, 2021). It is not immediately clear why the estimated returns differ across studies, as the empirical approaches use some combination of individual fixed effects and FE-cg. Differences in methods do not explain the differences in findings across studies.

Several papers also explore heterogeneity by gender, area of expertise, and program rank/quality. Arcidiacono, Cooley, and Hussey (2008) estimate returns to the MBA that are similar for men and women (4.8 percent and 3.9 percent, respectively). However, among both men and women there is a large return to program quality: the return to obtaining an MBA from a top-10 ranked program is 16.8 percent for women and 19.6 percent for men. Grove and Hussey (2014) estimate the return to MBA program quality using a quality index (Black and Smith, 2006) that consists of measures such as student characteristics (e.g., average GMAT and undergraduate GPA), faculty characteristics (e.g., publication count and percent with a Ph.D.), and program characteristics (e.g., rejection rate and class size). Using an individual fixed effects method, they find that the earnings of MBA degree recipients varies by program quality: earnings increase by 9.2 percent for each standard deviation increase in the program quality index.

A final dimension of heterogeneity is the area of focus within an MBA degree. Some MBAs focus on finance, while others focus on management or accounting. The only evidence on differential returns to MBA concentrations comes from Grove and Hussey (2011). They again use individual fixed effects models and estimate substantially higher returns to MBAs who concentrated in finance and management and information systems (MIS). The overall return to an MBA of 7 percent is half the return in these two concentrations. The return to other concentrations does not differ substantially from the mean.

A related literature has emerged that examines gender-specific returns to the MBA. Given the persistent gender earnings gap and the high aggregate return to an MBA, combined with historical under-representation of women in business fields, it is important to understand how the MBA differentially impacts male and female earnings. Arcidiacono, Cooley, and Hussey (2008) find little difference in gender-specific returns to an MBA. However, in an analysis using selection on observables approaches among GMAT test-takers, Montgomery and Powell (2003) find that earning an MBA is associated with a decline in the gender wage gap (among those who take the GMAT) from 14 to 9 percent. The difference in findings between this paper and Arcidiacono, Cooley, and Hussey (2008) is notable and could be driven by differential selection of women into earning an MBA based on secular earnings growth.

Bertrand, Goldin, and Katz (2010) delve further into the sources of the gender pay gap among MBA recipients. They study subsequent earnings among MBA students at the Chicago Booth School of Business (a top-ranked program). While men and women have similar earnings directly post-graduation, after 10-16 years the gender gap is 0.6 log points. Furthermore, 13 percent of women are not working 10 years post-MBA, while only one percent of men are out of the labor market. One of the strengths of their paper is the ability to examine the mechanisms behind the emerging pay gap. They decompose the change in the male-female earnings gap into the part due to MBA courses taken, time out of the labor market, and differences in weekly hours. Together, these factors explain 84 percent of the total pay gap 10-16 years post-MBA. Their results align with evidence that women, and in particular mothers, prefer jobs with lower and more flexible hours while high-earnings occupations tend to require longer hours (Goldin, 2014). The lack of flexibility in hours among MBA-related professions and the convex return to hours worked in those professions likely leads many women to choose lower-paying occupations for their non-pecuniary attributes. This paper is important in establishing that the structure of the labor market for MBA graduates leads to a large difference in male versus female earnings over time, even among graduates of an elite MBA program.

5.4.3. The Returns to a Law Degree

Estimates of the return to a law degree (JD) are much higher than the estimate for an MBA, as shown in Table 11, Panel B. The estimates range from 0.46 (Altonji and Zhong, 2021) to 0.61 (Simkovic and McIntyre, 2014) log points. These are very large earnings effects that are helpful in explaining why these students do not default at high rates on their student loans despite the large amount of debt they accumulate on average. The only gender-specific estimates come from Simkovic and McIntyre (2014), who use selection on observables methods (including controls for college major and the high school curriculum) combined with data from the Survey of Income and Program Participation (SIPP). They find estimates of 0.57 for men and 0.70 for women. That the earnings effects are larger for women is noticeably distinct from the MBA gender-specific analyses, despite most legal professions requiring long hours and exhibiting convex returns to hours worked. Together, these factors would depress earnings among women who hold a JD relative to men as they obtain experience. No papers exist that examine how earnings evolve among male versus female JD recipients and the role of non-pecuniary job characteristics in explaining any divergence over time; this is an interesting question for future work.

A core focus of the research on the return to a postsecondary law degree is on differential returns to program quality. The market for law school admissions is highly competitive, and there are large differences in the students, tuition, and salaries between top-ranked programs and lower-tier law schools. Oyer and Schaefer (2019) and Sander and Bambauer (2012) use survey data from After the JD (AJD) that allows them to measure post-JD earnings as well as respondents' background characteristics and some characteristics of their undergraduate major. Both find evidence of larger returns for top-ranked programs relative to lower-ranked programs. For example, Oyer and Schaefer (2019) find a return to a top-10 program is 0.12 log points higher than the return to a program ranked 11-20. The relative magnitudes in Sander and Bambauer (2012) are similar.

These papers are able to examine some of the mechanisms underlying this return to program quality. Sander and Bambauer (2012) show that the positive returns to law school quality are experienced predominantly by the higher-performing students: conditional on grades in law school, the return to program quality is much smaller. Some of this phenomenon can be explained by those who earn higher grades being more likely to be hired by larger firms out of graduate school and being more likely to be retained by these firms for up to 15 years. Oyer and Schaefer (2019) show as well that graduates of top-10 law schools are more likely to be placed in top prestigious firms, where earnings are higher. The ability to obtain a job with such firms is part of the higher returns students experience. They also document that undergraduate college quality and law school quality are substitutes: the return to elite law schools is accrued by those who do not come from elite undergraduate institutions. One explanation for this finding is that those who attend elite undergraduate universities have access to networks that help them obtain jobs at top firms, even when they go to lower-ranked law schools.

5.4.4. The Returns to a Medical Degree

Panel C of Table 11 presents estimates on the return to a medical doctoral degree (MD). There has been less work examining the return to medical school in the U.S. than on the other graduate degrees discussed above. Altonji and Zhu (2021) and Altonji and Zhong (2021) provide the only U.S.-based estimates, ranging from 0.62 to 0.76 log points. These large estimates align with the high pay for doctors documented by Gottlieb et al. (2020) and suggest that these high mean earnings reflect a large return to this degree. They also help explain how doctors can take out such large sums in student debt without experiencing high default rates.

The highest-quality studies of the return to medical degrees come from the Netherlands (Ketel et al., 2016; 2019). In the Netherlands, high school students apply directly to medical programs, and if admitted earn an undergraduate and a medical degree. Quotas on the number of students permitted to be admitted to Dutch medical schools lead to excess demand, which is resolved through randomized lotteries. These lotteries are weighted, such that those with a higher high school GPA are more likely to be admitted. Conditional on GPA (and on applying), admission to medical school is random, which allows them to overcome the identification problems associated with estimating the returns to graduate degrees more credibly than any prior study.

Ketel et al. (2016) use Dutch administrative data among applicants in the 1988-1999 admission cohorts that are linked to labor market outcomes. They find that doctors earn a 20 percent premium in every year after graduation, which leads to a 50 percent premium after 22 years. This estimate, while large, is somewhat smaller than the U.S. estimates discussed above. This could be because of lower wage variance in the Netherlands relative to the U.S., or it could indicate some residual selection bias in the U.S. studies. They estimate an internal rate of return of 36 percent, which is so high in part because medical school is heavily subsidized in the Netherlands. The authors argue that the quotas generate monopoly rents that accrue to Dutch doctors after ruling out mechanisms such as longer hours, more investments in human capital, or differences in family arrangements. Ketel et al. (2019) conduct a very similar study among Dutch dentists, who also are admitted to school using randomized lotteries. Their findings and conclusions closely mirror those for medical doctors.

As discussed in Section 5.4.2 with respect to MBA degrees, the structure of the labor market and return to hours worked can lead to large earnings differences by gender because men and women on average trade off pecuniary and non-pecuniary aspects of jobs differently as they age. Similar evidence exists in the medical profession. Sasser (2005) uses data from the Young Physician Survey to first document that women earn \$37,275 less than men (\$129,704 versus \$92,429, in 1990 dollars), work about nine hours less per week, and earn a lower hourly wage. This is consistent with a convex return to hours worked. Much of the earnings difference can be traced to family formation: women earn 11 percent less than men when they are married, 14 percent less when they have one child, and 22 percent less when they have more than one child. These results suggest that doctors who are mothers take jobs that require fewer hours and have lower pay, similar to the evidence from Bertrand, Goldin, and Katz (2010) among MBA recipients.

Chen and Chavalier (2012) show similar evidence comparing earnings of primary care physicians to physician assistants (PA). While male and female PAs earn less than MDs, the MD earnings premium is larger for men. However, women doctors work fewer hours than male doctors, and this difference in hours makes the return to an MD lower than the return to a PA degree for women. On average, female primary care physicians do not work enough hours to recoup the relatively higher cost of obtaining an MD. Men, on average, do work sufficient hours to recoup this relative cost. Like with MBA degrees, the market structure and the convex return to hours worked leads to a substantial gender pay gap.

6. Differential Returns and Inequality

The prior sections of this Chapter discussed evidence on the wide variation in the returns to different types of postsecondary investment. In particular, there is a large body of work using various identification strategies showing that students investing in four-year relative to two-year degrees, who attend more selective four-year universities, and who obtain degrees in more technical fields experience the highest returns. These differences beg the question of whether and how the postsecondary system in the U.S. contributes to aggregate increases in inequality. The contribution of higher education to inequality is a function of differences across groups in the return to education multiplied by differences across groups in the types of educational investments they make. The evidence discussed above suggests that economically disadvantaged students obtain similar returns to a given postsecondary investment as their higher-SES counterparts, sometimes even greater returns. Hence, the contribution of higher education to inequality rests predominantly on whether disadvantaged students are systematically investing in lower-return education choices and degrees.

In this section, we first discuss the evidence on differences in postsecondary investments made by students from higher versus lower-SES backgrounds. A multitude of studies document that students from lower-SES backgrounds are less likely to attend college, and conditional on attending are more likely to enroll in less-selective colleges with lower completion rates and lower returns. We next discuss the barriers faced by lower-SES students and the reasons why these gaps exist in postsecondary investment. We conclude the section by presenting research on different interventions that seek to close these gaps, focusing either on reducing specific barriers faced by lower-income students or on addressing a range of barriers faced by these students.

6.1. Evidence on SES Differences in Postsecondary Investments

A sizable body of work has arisen that documents the fact that students from lower-SES backgrounds invest less in postsecondary education and attend less-selective institutions. Figure 6 shows the distribution of initial postsecondary enrollments by students in the NLSY97 within two years of high school graduation, adapted from Lovenheim (2011). Panel A presents averages over the entire sample of the percent of students not enrolling, first enrolling in a two-year college, a non-flagship public four-year university, a flagship public university, and a private four-year university.⁶⁸ About two-thirds of students enroll in college, with 26 percent first enrolling in a two-year college. Only 6 percent of students attend a public flagship university, and 12 percent enroll in a private four-year college.

Panel B shows these patterns separately by family income in 1997. There are large differences by income both in the likelihood of attending college and in the types of colleges students attend. In the highest income group whose parents make over \$125,000 per year, 87 percent of students enroll, while 68 percent enroll in a four-year college. Sixteen percent enroll in a public flagship, and 22 percent enroll in a private university. The enrollment behavior of students from the lowest-income families earning under \$25,000 per year differ substantially: only 44 percent enroll in any college, 26 percent attend a two-year college, and 18 percent enroll in a four-year university. Only 2 percent of these low-income students attend a flagship public and 5 percent attend a private university. The returns to the types of colleges that lower-SES students attend are lower, as shown in Section 4, suggesting that the postsecondary system acts to reinforce pre-existing disadvantage.

These postsecondary enrollment patterns are reinforced by the findings from a number of studies. Pallais and Turner (2006) present enrollment percentages of low-income students from the most selective public and private colleges in the U.S. using data from the Institute for College Access and Success, Inc. Less than 10 percent of students from the top-30 ranked universities came from families with earnings under \$30,000. At five of these universities, fewer than 5 percent of students were from such families. Students whose parents earned between \$30,000 and \$60,000 represented 13 percent of enrollments at the top 30 universities. At state flagship universities, only 9 percent of students were from families with earnings below \$30,000. These results complement

⁶⁸ The sample is too small to show four-year private enrollment by selectivity, and for-profit enrollment was too low during this time period to separately examine.

those in Figure 6 by showing that students from low-income backgrounds are highly under-represented at the most elite universities in the U.S.

Hoxby and Avery (2013) compare the application behavior of high-achieving, low-income students to their high-income counterparts.⁶⁹ They show that over 40 percent of high-achieving, low-income students apply to non-selective colleges, even though if they did apply to more-selective schools they would get in and the generous financial aid at these institutions likely would make them less expensive to attend. Under 8 percent of high-income, high-achieving students apply to non-selective colleges. Hoxby and Avery (2013) measure the match quality of students and institutions using the difference between the student ACT/SAT and the institutional average. Among those who apply to selective universities, low-income, high-achieving students apply to far lower-match universities than do high-income, high-achieving students. Taken together, this evidence shows that even among high-achieving students there is a large selectivity gap between higher and lower SES students. Furthermore, they show that the low-income, high-achieving students who do apply to very selective colleges come from a small number of “feeder” schools from larger cities. This leads to what they term the “missing one-offs”: low-income, high-achieving students from high schools where they are unusually academically strong tend to not apply to selective colleges. Their findings suggest that knowledge about the postsecondary system, financial aid, and the value of attending a highly-selective university may be lacking among these students.

Chetty et al. (2020) use detailed data from U.S. tax records between 1999 and 2003 to examine income distributions and income segregation across colleges.⁷⁰ They focus on “Ivy-plus” universities, which are the eight Ivy League institutions (Harvard, Princeton, Cornell, Brown, Dartmouth, UPenn, Yale, Columbia) plus Duke, MIT, Stanford, and University of Chicago. These are the most elite and selective universities in the U.S. They document that 14.5 percent of students come from families in the top one percent of the income distribution, which is more than the 13.5 percent who come from the bottom half of the distribution. Under four percent of students come from families earning below \$25,000, which is the bottom quintile of the income distribution. The

⁶⁹ They define high-achieving students as those in 12th grade who have an ACT or SAT score of at least the 90th percentile and a high school GPA of at least an A-. Low-income and high-income students are those who are from families in the bottom and top quartiles of the income distribution, respectively.

⁷⁰ Specifically, they use a de-identified, publicly-available dataset that links federal tax data to data from the College Board, ACT, and the Department of Education. The data are available at <https://opportunityinsights.org/data/>.

pattern at other highly-selective private universities are similar though somewhat less dramatic. In the highly-selective public sector, 5.5 percent of students come from the bottom 20 percent, while 2.5 percent come from the top percentile. Hence, the elite public sector does not have much higher representation than the elite private sector among low-income students, but it does have lower representation among very high income students. The implication is that middle class students are more represented in the elite public sector.

In contrast, low-income students are much more likely to attend two-year and for-profit colleges than their high-income counterparts. Two-year colleges have almost 15 percent of students from families earning below \$25,000, and only 0.5 percent are from families in the top percentile. Over 20 percent of students in two-year and four-year for-profits are from the bottom quintile, while the share of top percentile students is similar to the share in two-year colleges. These tabulations highlight that there is a large amount of income segregation by parental income across college in the U.S. Indeed, Chetty et al. (2020) show that the income segregation across colleges is as large as the income segregation across neighborhoods. Students from high-income families are by far the most likely to enroll in the highest-return universities, suggesting that elite universities in America act as “bastions of privilege” rather than “engines of opportunity” (Pallais and Turner, 2006).

6.2. Explanations for SES-based Differences in Collegiate Investments

The studies discussed above use a range of different datasets that lead to the same conclusion: low-income students apply to and attend colleges very differently than do students from higher-income families. This is the case even when focusing on high-achieving students. Chetty et al. (2020) calculate that equalizing college quality across the family income distribution would reduce the intergenerational persistence of income by 27 percent among college students.

The large and persistent SES gaps in collegiate investments has led to a large body of research that examines policy interventions aimed at supporting postsecondary investments among students from low-income backgrounds. In order to interpret the evidence from these studies, it is important to consider the various reasons why these SES gaps exist. What barriers do low-income students face when they seek to invest in a postsecondary education? The answers to this question inform where to target policies to reduce these barriers.

We follow Andrews, Imberman, and Lovenheim (2020) in dividing barriers faced by low-income students into five groups. First, students from families with lower incomes at the time of

the collegiate investment decision likely had fewer resources throughout the child's life. This leads to what Carneiro and Heckman (2002) term "long run credit constraints." These credit constraints reduce the amount parents are able to invest in their children's education throughout their childhood, leading to lower academic preparation for college by the time these students reach the end of high school. As shown in Figure 2, students with higher academic preparation for college are much more likely to be successful in college. Furthermore, SES-based differences in college preparation will lead higher-SES students to be more competitive in the application process for elite schools (e.g., Goodman, Gurantz, and Smith, 2020).

Pallais and Turner (2006) and Chetty et al. (2020) show that long run credit constraints as reflected in student academic preparation for college are empirically relevant. Both papers demonstrate that there are relatively few students from low-income backgrounds who are academically eligible (without the exercise of admission preferences) to attend a highly-selective college. Chetty et al. (2020) simulate the effect of equalizing application, admission, and matriculation rates across the family income distribution, conditional on test scores. They show that this would have little effect on representation by low-income students at elite colleges, since the supply of high test score students from low-SES backgrounds is low.⁷¹ The policy implications of these results are important, as they indicate that the SES differences discussed above are driven by more than "access." Substantially reducing SES gaps either requires highly-selective universities to practice admissions preferences for low-income students or K-12 or other youth-focused policies that will enable students from disadvantaged backgrounds to obtain sufficient pre-collegiate training. Such policies are of high importance but are beyond the scope of this Chapter, and we do not discuss them below.

Second, there are differences in the information that high- versus low-income students have about college. This is evident in the patterns of college applications in Hoxby and Avery (2013). That such a large percentage of high-achieving, low-income students do not apply to any selective colleges and among those who do are more likely to apply to institutions where they are academically undermatched suggests that they are unaware of optimal (i.e., selectivity-maximizing) application strategies, the enhanced financial aid for which they would be eligible at elite colleges, and the returns to enrolling in more-selective schools. This insight has led to a large

⁷¹ This would reduce cross-college income segregation substantially, however, because middle-income students would be much more likely to attend elite colleges.

literature that seeks to overcome these information barriers and to directly recruit students that we discuss below.

Third, because high-income students are dramatically over-represented in elite postsecondary institutions, students from other backgrounds may experience both academic and social “mismatch.” Even conditional on high school GPA, lower-SES students are likely to be less-academically prepared for college. Because of income segregation across neighborhoods, on average students from lower-income backgrounds attend lower-quality K-12 schools. This is a dimension of long-run credit constraints, but it is conceptually distinct in that it leads even these high-achieving students to be less-prepared for college than students from wealthy backgrounds. Similarly, wealthy students form the dominant cultural majority at elite postsecondary schools, which can lead lower-income students to feel that they do not belong. Cultural and academic mismatch can induce students from disadvantaged background to not apply at all or to drop out/transfer early in their college careers.

Fourth, the complexity of the financial aid and application processes may dissuade many students from participating. Applying to a selective college requires a lot of knowledge and effort on the part of the applicant. There are various deadlines and application requirements that vary across schools. Students from lower-income families likely have less access to guidance counselors and adults with experience in the selective admission process. As detailed by Dynarski and Scott-Clayton (2006) and Page and Scott-Clayton (2016), the financial aid process and college application and enrollment process is extraordinarily complex and time-consuming, which can create a high barrier for access.

The incidence of the complexity costs in applications and financial aid are likely to fall on the most disadvantaged students, thus exacerbating SES gaps in college investment. While the costs of the complexity in these systems are small relative to the payoff of attending an elite college, there is much evidence from the behavioral economics literature showing that small costs can have large impacts on behavior. Especially because most college applicants are teenagers, behavioral factors like time inconsistency, loss aversion, identity salience, and default behavior can deter many students from engaging with a complex admission and financial aid system. French and Oreopoulos (2017) provide a review of the literature on simplifying the college admission process. They argue that the results fit a model wherein students are inattentive to college possibilities, miss deadlines, and simply do not take advantage of possible collegiate opportunities.

We discuss some of the research below on complexity reduction and reducing the costs of engaging with the application and financial aid systems. Many of these interventions are low-cost, making even small positive effects highly cost-effective.

Fifth, there could be short-run credit constraints (Carneiro and Heckman, 2002) that make it impossible for students to secure the resources to finance their preferred college education at a sufficiently low interest rate to make the investment worthwhile. The more generous financial aid available at more-selective colleges makes this an unlikely barrier, however many low-income students likely do not know that more-selective schools will be cheaper for them to attend than less-selective schools. So, there could be effective credit constraints that bind because of a lack of information about financial aid policies. Lovenheim and Reynolds (2013) provide suggestive evidence of this mechanism. They show that when families experience a housing wealth increase when they have a child nearing college age, the child is more likely to attend a more-selective college. The mechanism driving this effect is students being more likely to apply to state flagship universities. Since financial aid (and thus net cost) is unknown at the time of application, this result suggests that students respond to perceived costs and ability to pay when making application decisions.

In the remainder of this section, we will discuss several policies and interventions that have been aimed at reducing these barriers to success among low-income students in higher education. We focus specifically on five types of policies: 1) recruiting and information simplification; 2) lowering college entrance exam costs; 3) coaching and mentoring high school students; 4) promise programs; and 5) comprehensive supports that address multiple barriers.

6.3. Recruiting and Information Simplification

A number of different randomized experiments have sought to overcome the information barriers with some combination of directly recruiting high-achieving, low-income high school students, providing them with information about the application and financial aid process as well as college options, and simplifying the application system. Hoxby and Turner (2013) reports the results of one of the first such randomized controlled trials. The authors implemented a nationwide experiment, called the Expanding College Opportunities Project (ECO). The experiment had many treatment arms, and like the authors we focus on the aggregated intervention.

Treated students received a range of information and application supports. They received application guidance, attempting to mimic the kind of advice about the range of schools to which

they should apply and financial aid guidance that they would receive from a high-quality guidance counselor. They also are given information on graduation rates of the nearest colleges, the state public flagship university, other in-state selective colleges, and a select number of out-of-state selective colleges. They additionally receive extensive information on net costs, nationally and locally, for three low-to-moderate income groups, a financial aid guide, a list of colleges that satisfy 100 percent need, and application fee waivers. This is a randomized controlled trial focusing on high-achieving, low-income students, defined as those in the bottom tercile of the income distribution for families with a high school senior and who score in the top decile on the ACT or SAT. Taken together, the intervention is aimed at reducing the information asymmetries faced by low-income, high-achieving students by providing clear and semi-personalized information about net costs they are likely to face, optimal application strategies, and local and national college options. They also help subsidize more applications through fee waivers.

Using data from the National Student Clearinghouse linked to their sample, they find that treated students apply to more schools and apply to more-selective schools. Furthermore, treated students were admitted to more colleges and they were far more likely to be admitted to a more-selective and higher-resources college (because they applied to them). They are more likely to enroll in such institutions as a result. The evidence in Hoxby and Turner (2013) shows that high-quality information sent to students on application strategies, college options, and net costs can have large impacts on student application and enrollment behavior. Notably, the intervention only cost about \$6 per student. While it is too early to estimate long-run returns, the effects on college quality suggest that this is a cost-effective intervention.

A related randomized controlled trial was conducted by the University of Michigan that sought to eliminate uncertainty about financial aid for low-income, high-achieving students in the state (Dynarski et al., 2021). The “HAIL Scholarship” gave guaranteed no tuition and fees to low-income, high-achieving students in Michigan if they were admitted to and enrolled in the University of Michigan. Critically, the amount of aid given to students did not increase. Rather, the salience of the aid increased. They find large effects of the aid guarantee: enrollment more than doubled (from 26 percent to 68 percent), and over 25 percent of the treated students would otherwise not have enrolled in any college. The intervention did not simply shift students across selective universities. Making aid salient increased college-going and college quality substantially, which highlights student aid salience as a core hurdle among low-income, high-achieving students.

The large enrollment effects in Hoxby and Turner (2013) and Dynarski et al. (2021) have not been found in lighter-touch recruiting experiments. Hyman (2020) mailed letters to 50,000 high-achieving high school seniors in Michigan that encourages them to consider attending college and points them to an informational website. There is no overall effect on enrollment, which likely is related to the very low take up rate of the information, however economically disadvantaged students were about 1.4 percentage points more likely to enroll in college as a result of receiving this letter. At a cost of 15 cents per letter, even small effects on this subset of the population could be cost-effective. Gurantz et al. (2021) report on the results of an information experiment run by the College Board, wherein low- and middle-income students were mailed brochures and sent informational emails encouraging them to apply to college and supplying information about the application process. In a large RCT with 785,000 participants, they find no effect on enrollment. One explanation for why these studies find null or small effects is that they did not provide clarity on college costs or simplify the financial aid process.

6.4. Lowering College Entrance Exam Costs

College entrance exams form another potential hurdle in the higher education system for students from low-income backgrounds. Not only do these students lack the resources to engage in extensive exam preparation, but they may also not even take an entrance exam. This closes off the opportunity to enroll in many selective colleges. Furthermore, many students who take these exams are dissuaded from applying to a broad set of schools by the costs of sending more than a small number of free scores to colleges. A growing literature examines the effects of policies aimed at expanding test-taking and reducing the cost of score-sending.

Bulman (2015) constructs a panel of all SAT test-taking centers in the U.S. linked to student enrollment outcomes from the National Student Clearinghouse. He exploits variation in test availability from the opening and closing of test centers as well as district policies on free in-school testing and default registration. He finds evidence that expanding access to the SAT through the opening of a test center in one's school (or in one's neighboring school) as well as policies aimed at reducing the cost of taking the exam led more students to take the test. Those induced to take the SAT are much more likely to enroll in a four-year college, and they also are just as likely to graduate as inframarginal attendees. These findings highlight test-taking costs as an important barrier to postsecondary investment by many students.

Goodman (2016), Hyman (2017) and Hurwitz et al. (2015) examine the effect on college attendance of policies that mandate all students take a college entrance exam. Goodman (2016) examines the adoption in five U.S. states of mandates requiring high school juniors to take a college entrance examination. She shows the mandates led to large increases in test-taking and to a 20 percent increase in selective college enrollment. Hyman (2017) uses the rollout of free, mandatory ACT testing for high school students in 11 states and shows resulting increases in four-year college enrollment. The effects are largest among those from the lowest-income high schools. Hurwitz et al. (2015) show that a similar policy in Maine for the SAT increased college enrollment.

Both the SAT and ACT allow students to send up to four free reports to colleges, after which students must pay \$16 (ACT) or \$12 (SAT) per additional score sent. Despite the relatively low marginal cost of sending additional scores, many students only send the free reports. Pallais (2015) shows that when the ACT increased the number of free reports students could send from three to four in 1997, students applied to more colleges, applied to a wider set of schools, and among low-income students there was an increase in the quality of colleges in which students enrolled. Hurwitz et al. (2017) study a related policy change by the College Board for the SAT. In 2007, they allowed low-income students who received a fee waiver for the exam to send an additional four scores any time during their high school career. This subsidy, worth at most \$48, increased the likelihood of sending eight or more scores, led to a higher four-year college attendance rate, and increased the six-year BA completion rate by 2 percentage points.⁷²

The sensitivity of entrance exam taking and college enrollment to changes in the cost of test-taking and score sending is somewhat surprising. These ultimately are small costs relative to the substantial returns to collegiate investment documented in Sections 4 and 5. The results from these studies are consistent with behavioral models such as time inconsistency and default behavior bias that can amplify the effect of even small costs to the application process. That the incidence of these costs tends to fall more on disadvantaged students can increase inequality in higher education. Policies that lower the costs of test-taking and score sending, including mandating taking college entrance exams, can have large impacts on college application behavior and collegiate attainment, especially for students from low-SES backgrounds.

⁷² Goodman, Gurantz, and Smith (2020) find large effect of retaking the SAT as well. Using the fact that students are more likely to retake the exam when they are just below multiple of 100 (left digit bias), they show that retaking the SAT leads to higher scores and subsequent college enrollment. The effects are particularly large among low-income students, who otherwise are much less likely to retake the SAT than their higher-income counterparts.

6.5. Coaching and Mentoring High School Students

One direct way to overcome information barriers on the path to higher education is through coaching and mentoring of high school students by those knowledgeable about the system. A number of recent studies have run randomized controlled trials to identify the effect of counseling and mentorship on college enrollment. The details differ across interventions, as do the costs, however the majority of studies find that coaching and mentoring positively affects college-going.

Carrell and Sacerdote (2017) run a series of randomized controlled trials that assess the role of information and college coaching on college enrollment in New Hampshire. They match students interested in attending college but who have not taken steps to apply (as identified by their guidance counselor) with Dartmouth College students. The Dartmouth students provide direct application assistance as well as help filling out the FAFSA, and treated students had all application fees paid. Simply providing information about college and financial aid had no effect on enrollment, which is consistent with the null effects of information-only experiments discussed in Section 6.3. Mentoring had a large, positive effect on college enrollment among women, with smaller effects among men that were not statistically significantly different from zero. Treated women also were more likely to persist. Examining mechanisms, they show evidence that the treatment effect was smaller among men because they expect a higher wage return to only a high school diploma. They further present evidence that the intervention works by providing the direct supports that were largely absent from parents and teachers. While intensive, mentoring programs can reduce gaps in college enrollment supports that widen SES-based differences in collegiate investment.

A range of college counseling studies have found results that are broadly aligned with those in Carrell and Sacerdote (2017), with the exception that they often show larger impacts on men.⁷³ Oreopoulos and Ford (2019) report on an RCT for the LifeAfterHighSchool program in Ontario, Canada, which focused on high schools with low college enrollment rates. Students attended three workshops in which they picked programs of interest, applied to them, and completed financial aid applications. They find large increases in college application rates for both men and women. Barr and Castleman (2018) examine a multi-cohort and multi-site RCT of the Bottom Line

⁷³ Bettinger and Baker (2014) and Oreopoulos and Petronijevic (2018) show evidence of positive effects of counseling interventions among students already in college on collegiate attainment. They generally find that more student support by college counselors leads to better student outcomes. Barr et al. (2021) find no effects of counseling and information provision on postsecondary decisions made by U.S. Army veterans, however.

program. This program provides college advising to low-income students, helps them select appropriate, high-quality programs, and assists with applications and financial aid. They find treated students attend higher-quality four-year colleges and are more likely to persist. They also present direct evidence of the importance of student-counselor interactions, which aligns with the main mechanisms in Carrell and Sacerdote (2017). Barr and Castleman (2021) show that this program additionally has large, positive effects on BA attainment rates that are driven by the increased enrollment in higher-quality colleges.

Bettinger and Evans (2019) address the role of within-school spillovers by running a whole-school RCT of college counseling. They employ recent college graduates as counselors, who help students with college applications and the college transition process. While there are no overall effects on enrollment, low-income and Hispanic students are more likely to enroll in college. In contrast to the other college coaching studies, the enrollment effects are concentrated in two-year colleges. This finding could be a result of focusing on all students, rather than just high-achieving or college-oriented students.

Advances in communication technology, which have expanded dramatically due to the COVID-19 pandemic, offer the opportunity to expand advising services to a larger set of students at a lower cost. Gurantz et al. (2020) provide evidence of the potential effectiveness of virtual advising. They study an RCT that assigned advisors from the College Advising Corps to high-achieving students, who met virtually with students to help coach them through the application process. Receiving virtual advising increased the quality of four-year colleges students attended but did not increase enrollment overall. Notably, this RCT was done prior to the recent COVID-19 pandemic. Sullivan et al. (2021) examine a randomized, multi-cohort virtual advising interventional called CollegePoint. They find that virtual advising leads to a modest increase in college quality, with a larger effect during the COVID-19 pandemic. These findings suggest that recent expansions of video technology to communicate will increase the scope to use virtual college advisors to help low-income students apply to college.

6.6. Promise Programs

In 2005, a group of anonymous donors in Kalamazoo, MI pledged that they would pay 100 percent of tuition and fees at a public Michigan college for any graduate of Kalamazoo Public Schools as long as they were continually enrolled in the district since 9th grade and currently live in the district. The “Kalamazoo Promise” gave rise to a number of similar “promise” programs in

different cities across the U.S. and has spurred a growing movement for tuition-free college. Promise programs are place-based, meaning that they attach financial aid to where students graduate from high school, and they typically cover tuition and fees at either four- or two-year in-state public institutions. Promise programs tend to be located in school districts serving low-income (typically urban) students. These programs are similar in many ways to the HAIL Scholarship discussed in Section 6.3 (Dynarski et al., 2021), greatly simplifying the financial aid process for qualifying students. A main difference is that promise programs are place-based rather than income-based, which raises important questions about how well these policies are targeted and whether they spur gentrification.

Bartik, Hershbein, and Lachowska (2021) provide some of the most comprehensive evidence on the effects of promise programs on postsecondary outcomes. They study the Kalamazoo Promise using two strategies. The first is a within-district approach that compares changes in student outcomes among eligible versus ineligible students across cohorts when the program was implemented. Students typically are ineligible because they have not lived in the district for a sufficient amount of time. While ineligible students are a sensible control group, it is possible that late entrants to the district are changing relative to other students around 2011. They thus supplement their results with a cross-district difference-in-difference analysis surrounding program implementation. They find that four-year college enrollment increased by 18-27 percent due to the promise scholarship, with particularly large enrollment gains in the top public colleges in the state. Students were also more likely to complete college, with degree attainment rising by 12 percentage points. As with the HAIL scholarship, student aid guarantees lead to large increases in postsecondary investments, even without substantial changes in eligibility for aid.

The Pittsburgh Promise was a similar program launched in 2008. To obtain the full scholarship, students had to be enrolled in the Pittsburgh Public Schools since kindergarten and have a high school GPA above 2.5. Using both an RD surrounding the GPA threshold and a difference-in-difference design comparing eligible to ineligible students when the program was implemented, Page et al. (2019) find increases in overall and four-year college attendance. The program does shift some students to attend in-state colleges who would have enrolled in out-of-state colleges, however. There also is a sizable increase in the likelihood that treated students persist into their second year.

Twenty-three states have launched promise programs for community colleges, essentially making public two-year college tuition-free for in-state residents. These are “last dollar” programs, meaning that they pay the difference between tuition and fees and the federal, state, and institutional aid for which students are eligible. Unlike the Kalamazoo and Pittsburgh promises, the restriction to community colleges could divert some students from the four-year to two-year colleges, and so the overall effect on collegiate attainment is ambiguous. Furthermore, students have to fill out the FAFSA and apply for financial aid, so these policies do not address this barrier to entry. Two papers study these community college promise programs: Gurantz (2019) in Oregon and Carruthers, Fox, and Jepsen (2020) in Tennessee. Gurantz (2019) uses College Board data that links PSAT and SAT-taking students to National Student Clearinghouse data. Oregon subsidizes the PSAT, which leads to high test-taking rates. Using six control states with similar policies as a control group in a difference-in-difference approach, Gurantz (2019) finds a 4-5 percentage point increase in college enrollment in the first two years of the program. In the first year, the effect comes from diversion of students from four- to two-year colleges. In the second year, the effect is driven more by students who otherwise would not attend college.

Carruthers, Fox, and Jepsen (2020) study the effect of the Knox Achieves promise program, which promised tuition-free community college for all graduates in Knox County, Tennessee. This program was later expanded to all students in the state. They are uniquely able to examine earnings as well, which is important because of concerns over whether students are harmed through diversion from four-year to two-year colleges. To identify program effects, they use difference-in-difference methods across school districts that leverages the introduction of the Knox Achieves program as well as matching approaches that pair each Knox Achieves student with observationally-similar students outside of Knox County or prior to the program introduction who thus were ineligible for the program. Overall, the Knox Achieves program raised the number of postsecondary credits students accumulate as well as certificate and associate degree attainment. Bachelor’s degree attainment rates decrease slightly. Focusing on low-income students, community college attainment rises while BA completion decreases. This is evidence of diversion among this population. Nonetheless, earnings seven years after high school substantially increases, suggesting that any adverse effects of diversion are smaller than the increases in postsecondary attainment among students who otherwise would not have attended college.

6.7. Comprehensive Support Programs

The policies discussed above focus on addressing one dimension of the multiple disadvantages faced by students in the higher education system. Reducing one type of disadvantage may not be sufficient, however, if there are interactions among them. For example, a recruiting program can induce a low-income student to attend a more-selective college, but once they are there, they then face disadvantages from academic and cultural mismatch that can impede their success. The recognition that students from low-SES backgrounds face multiple disadvantages that need to be addressed simultaneously has led to a series of comprehensive interventions that seek to provide broad support for these students. These programs have the benefit of being comprehensive, but it then is challenging to disentangle the underlying mechanisms that can inform how we should best structure comprehensive support systems. Comparisons across programs with different features provides insight into optimal program design.

One set of studies on comprehensive support systems focuses on four-year, selective colleges. Angrist, Lang, and Oreopoulos (2009) report the results of an RCT at a satellite campus of a large Canadian University. Some students were offered support services such as mentoring by upper-class students and additional instruction. Others were offered financial incentives for obtaining a GPA above a certain level, and a third group was offered both interventions. They find large gender differences in effects. Women are more likely to use the advising services, and the use of support services is much larger when paired with financial incentives for earning a higher GPA. Academic performance of women increases from all treatments, but the effects are largest for those randomized into the dual treatment arm. Men, however, are not very likely to use services and are unaffected by any of the treatments.

Page et al. (2019) study the Dell Scholars Program, which was launched in 2004 and provides financial and non-financial resources to low-income college students. Recipients are given a total of \$20,000 in scholarship funds as well as extensive academic supports throughout their college enrollment. These supports are done through a web-based platform that allows Dell Scholars Program retention officers to flag potential problems and intervene with students who are struggling. Using review criteria cutoffs for scholarship receipt, Page et al. (2019) show large effects of receiving the Dell Scholarship on retention, academic performance, and Scholarship receipt. Clotfelter, Hemelt, and Ladd (2018) study a similar program at University of North Carolina – Chapel Hill, called the “Carolina Covenant.” This program provided financial aid and

academic support services to low-income students who attend UNC. Using both RD difference-in-difference techniques, they find increases in student performance but only after the student service supports were added to the program. Angrist et al. (2017) show similar results from an RCT of the Susan Thompson Buffett Foundation scholarship in Nebraska. The scholarship provides generous support of upwards of \$70,000 to study at a public university in Nebraska. Recipients enrolling in an in-state college also can participate in Learning Communities, which provides a suite of academic and support services to students. Their results show that scholarship recipients are more likely to be enrolled in college four years later and shifted many students from two- to four-year colleges. However, on-time degree receipt falls slightly among the treated group.

The programs discussed above do not have a recruiting component – they take student college selection as given. Given the lower propensity of low-income, high-achieving students to attend selective universities, pairing these supports with active recruiting may be particularly effective. Andrews, Imberman, and Lovenheim (2020) study the Longhorn Opportunity Scholars program in Texas that contained all of these components. The program began in 1999 and targeted Texas public high schools that served low-income students in urban areas that historically sent few students to University of Texas at Austin (UT-Austin). There is an active recruiting component to help induce high-achieving students from these schools to apply to UT-Austin, students receive \$5,000 per year in financial aid, and once enrolled students receive extensive academic support services that include special classes, dorm space, and enhanced advising. Andrews, Imberman, and Lovenheim (2020) use administrative data on all public K-12 students in Texas matched to all public higher education students in Texas and to quarterly earnings records from the UI system. They employ a difference-in-difference design that identifies observationally-similar untreated schools as a control group and show that the program increased enrollment in and graduation from UT-Austin. There is no overall graduation effect, but the fact that students are more likely to obtain a BA from the state flagship could lead to earnings gains. They find that earnings of students from treated high schools increased by 4.6 percent 12 or more years after high school, which corresponds to an 82 percent increase in earnings among those who attend UT-Austin.

A number of comprehensive support programs also have started at public two-year colleges. Since two-year colleges tend to enroll more economically disadvantaged students, comprehensive supports may be particularly effective in this setting. The most well-known of these programs is the City University of New York (CUNY) Accelerated Study in Associate Programs

(ASAP). This is a three-year program, launched in 2007, aimed at increasing AA attainment rates. The program had multiple components: students were required to attend full-time, they were assigned to a dedicated academic advisor with whom they were required to meet twice per month, they had to meet with a career and employment specialist once per semester, students received full “last dollar” tuition and fee financial aid as well as free metro cards, and the program offered blocked courses and consolidated schedules to ensure students could enroll in needed courses and maintain convenient schedules. Weiss et al. (2019) study the RCT component of the CUNY ASAP program and find an 18 percentage point increase in 3-year AA attainment.

Finally, Evans et al. (2020) study the effect of a “case management” program called Stay the Course (STC) in Fort Worth, Texas. This was an RCT designed to help low-income students at community colleges overcome the life challenges that can impede their progress towards a degree. Treated students were assigned a trained social service provider who offers coaching, mentoring, and referral services to other social services students may need. Students also have access to emergency financial assistance to help address adverse financial shocks they experience. Some students only received emergency financial assistance. Women in the full treatment arm were 7.4 percentage points (47.1 percent) more likely to obtain an AA degree. There were no detectable effects on men nor among those who only received emergency financial assistance, although the estimates are rather imprecise. Their findings indicate that the comprehensive support services are particularly effective for women in this population.

7. Conclusion

Postsecondary education in the U.S. and around the world consists of a large variety of different options, leading to a wide range of potential returns for students. The earlier literature and reviews focused on average returns to schooling. In this Chapter, we discussed how the returns to college differ by institution, type, program, credential, and coursework. Indeed, the returns differ on all of these dimensions, which are too numerous to recount parsimoniously. However, we are able to draw some broad conclusions. First, most studies find positive educational and earnings returns to college quality, measured either using institutional resources/student composition or the level of the institution. The magnitude of the estimates range from quite modest to substantial, depending on the specific context and the specific measure of college quality being used. Second, there are differential returns to fields of study, which are equally as large as the

variation in returns across the college quality distribution. The combination of the first two broad conclusions imply that the choice of college and major are of immense consequence to students, colleges, and policy makers. Third, the importance of college major and institution type are evident around the world, not just in the U.S. The country and corresponding higher education system under consideration determine the process by which students choose colleges and majors, which also informs the identification strategy being used, however the results in international settings tend to align with each other and with findings from the U.S. Fourth, even within a college or field of study, there are varied potential educational experiences, ranging from coursework to credentials sought, and they all contribute to longer-term outcomes.

Finally, we summarize the emerging evidence that students from low-SES backgrounds, even high-achieving students from these backgrounds, are less likely to enroll in college and when they do so are more likely to enroll in colleges and degree programs that have lower returns. This suggests that, on net, colleges and universities may act as bastions of privilege rather than as engines of opportunity. We discuss U.S. based efforts to remedy these inequities, some of which have proven effective at changing students' educational decisions and outcomes.

Beyond these broad conclusions on the returns to varied postsecondary investments, we stressed throughout the Chapter the variety of methods used by researchers that require different assumptions in order to interpret the estimates as causal. Clearly, some assumptions are stronger than others. We present a model that can be used alongside most papers in the literature to better understand the assumptions required of any estimator, regardless of whether they are explicitly stated in the study. All of the studies we discuss add value to the literature, and clarifying the assumptions required for identification will help researchers to better understand and state those assumptions as well as continue to relax them as data and methods progress.

Throughout the Chapter, we highlight open questions in the literature that we hope researchers use to guide their work. Although the open questions and areas are too numerous to recount in full, below we list some of the broad topics we believe are worthy of more research:

- While we focus on differential effects of various educational investments, these investments are likely to have heterogeneous effects on different types of students. Who accrues the returns and why is an open question. Related, what are the distributional effects of educational investments? Most studies only focus on mean effects, with some notable exceptions in recent years.

- We discuss the literature on the labor market and educational returns to different postsecondary options because this has been the core focus of prior research. However, the returns to education extend beyond the labor market and include key outcomes such as health, civic engagement, crime, family formation, and intergenerational effects. Understanding more fully how different postsecondary investments affect these outcomes is an important direction for future research.
- What are the returns to partially-completed schooling? Very few studies address this question, and finding exogenous variation in dropout behavior remains a challenge. The same issue arises when examining part-time versus full-time enrollment and the effect of transferring.
- What are the life-cycle returns to educational investments? We briefly discussed this in relation to the returns to major choice, but most studies use an outcome at a single point in time. Any returns have the potential to evolve over the life-course, and understanding the evolution of postsecondary returns is critically important for characterizing the returns to a given postsecondary investment.
- The research on returns to majors and graduate school, especially in the U.S., is yet to adopt the same identification strategies used to study the return to other educational investments. Most notably, the use of credible regression discontinuity designs is far less prevalent in the U.S., mostly because of the institutional setting. We suspect advances in data access will make this an important area of future work.
- The research we discussed does not pay much attention to labor supply and demand factors in the labor markets in which students will work. The returns students experience are likely to vary with both labor supply and demand. More work explicitly linking the return to skill with the supply of and demand for skill in a labor market is of high value.
- While there is much evidence on the return to majors and the return to enrolling in different types of postsecondary institution types and degree programs, there is little understanding of the return to specific coursework and how this varies across institution type. As better data become available, we expect more research to emerge that examines the return to enrolling in specific courses and shows how such returns vary across institutions.
- Higher education is changing rapidly and the returns to new methods of learning and schooling need to be assessed. These include virtual learning, different types of programs, hybrid two-year and four-year colleges, etc.

- Many studies consider the impact of an educational investment but do not calculate the return on the investment (ROI). The ROI is critical for comparing educational options, and we expect the future literature to focus more on this measure relative to the far more prevalent examination of gross returns. The studies that do calculate a ROI often are done by assuming some level of tuition and costs. As better data on college costs emerge, they will allow for more precise ROI estimates.

Estimating the returns to different types of educational investments has been an active area of research over the past few decades. Knowledge of the variation in returns is critical for the design of higher education policies around the world. The data and methods brought to bear are very different today than they were at the beginning of those few decades. Given the importance of postsecondary education to the labor market and the increasing return to skill faced by workers, this will remain an active area of research as data and methods continue to progress.

Table 2. U.S. Higher Education Statistics – By Selectivity and Level

	Four-year institutions*				Two-year institutions
	All	More selective	Selective	Inclusive	
Number of institutions					
Degree-granting institutions	2,443	427	684	1,332	1,314
Public institutions	756	119	262	375	880
Private non-profit institutions	1,376	300	405	671	86
Private for-profit institutions	311	8	17	286	348
Fall enrollment					
Total fall enrollment	13,826,347	4,461,451	4,420,299	4,944,597	5,812,533
Total fall undergraduate enrollment	10,868,777	3,258,657	3,543,443	4,066,677	5,812,527
Total fall undergraduate, degree-seeking enrollment	10,141,838	3,172,610	3,325,179	3,644,049	4,549,901
Total fall undergraduate, degree-seeking, first-time enrollment	1,898,563	695,958	626,078	576,527	987,804
Student characteristics					
Percent receiving financial aid	89.7	85.3	94.4	88.5	82.1
Percent receiving Pell grant	46.6	25.8	40.7	57.7	59.4
Percent URM**	33.7	20.6	26.5	42.7	42.1
Admissions					
Reporting institutions	1,714	424	661	629	98
Applications per institution	6,709	14,301	5,559	2,800	-
Admit rate	68.2	53.6	72.3	73.7	-
Average SAT scores	1,129	1,282	1,103	1,011	-
Tuition and fees					
Average published tuition & fees (in-state for publics)	\$21,127	\$34,922	\$22,848	\$16,651	\$8,526
Privates only: published tuition & fees	\$27,004	\$43,803	\$31,059	\$20,145	\$15,226
Publics only: published in-state tuition & fees	\$8,580	\$11,368	\$9,842	\$7,626	\$4,896
Publics only: published out-of-state tuition & fees	\$18,885	\$28,710	\$20,449	\$15,704	\$8,485
Expenses per student					
Instructional expenses per FTE	\$10,592	\$18,417	\$9,633	\$8,481	\$6,214
Research expenses per FTE	\$1,606	\$5,134	\$890	\$802	\$22
Public service expenses per FTE	\$640	\$998	\$626	\$527	\$202
Academic support expenses per FTE	\$2,975	\$5,664	\$2,385	\$2,384	\$1,457
Student service expenses per FTE	\$3,765	\$5,699	\$3,820	\$3,089	\$1,998
Institutional support expenses per FTE	\$5,760	\$8,030	\$4,689	\$5,561	\$3,235
All other core expenses per FTE	\$1,924	\$2,183	\$1,660	\$1,976	\$2,566
Graduation rates					
Graduation rate within 150% of normal time***	51.9	75.4	56.7	40.3	39.1

Source: U.S. Department of Education, National Center for Education Statistics, Integrated Postsecondary Education Data System (IPEDS), 2019. Retrieved from <https://nces.ed.gov/ipeds/use-the-data> on October 1, 2021.

*Four-year selectivity defined by Carnegie Classification.

**URM includes Black or African American, Hispanic, American Indian or Alaska Native, and Native Hawaiian or Other Pacific Islander.

***Includes the 2013 entry cohort for 4-year institutions and the 2016 entry cohort for 2-year institutions.

Table 3. Select Features of International Higher Education Systems

Country	System	Institution Types	Tuition	Application Process	Admissions Criteria	Exogenous Variation Used by Researchers	Papers
Australia	Decentralized	Mostly public universities (37/43). Universities and vocational.	Expensive and depends on degree, not charged upfront, debt repayment tied to income.	Mostly state-specific application, list preferences for college-majors.	State-specific high school exam scores. Can be admitted to multiple college-majors.	Admission exam cutoff.	Carroll, Heaton, & Tani (2014)
Brazil	Decentralized	Mostly public, some private. One-third students in for-profit. Three tiers.	Free at publics.	Apply to college and major.	High school exam for some colleges (ENEM); College-specific admission exam for some colleges (Vestibular).	Race-based affirmative action quotas.	Francis & Tannuri-Pianto (2012)
Canada	Decentralized	Universities (research, undergrad, colleges) and vocational institutions; mostly public.	Modest	Apply to each college separately.	Mostly high school GPA, often cutoff. Selective universities require additional info (recommendations, extracurriculars, etc.)		Boothby & Drewes (2006); Bar-Or et al. (1995)
Chile	Centralized	CRUCH colleges (public, private, private-parochial, all not-for-profit); Non-CRUCH colleges (mostly private, service lower-scoring students).	College-major specific. Free for many colleges for poorest 50% since 2016.	Rank up to 10 college-major in single application (CRUCH and most private colleges).	National exam score (PAA/PSU) and sometimes high school GPA. Admitted to one college-major.	Admission exam cutoff, student loan eligibility cutoff.	Hastings, Neilson, & Zimmerman (2013); Rodríguez, Urzúa, & Reyes (2016); Bucarey et al. (2020)
China	Centralized	First-tier (top ~100 are elite), provincial and local 4-years, vocational.	Low fees, often cheaper at elite colleges.	Rank up to 5 colleges in single application, with 3-5 majors in each college. Knows min. admission cutoffs.	National exam score (Gaokao); provincial quotas; affirmative action for ethnic minorities, child of military casualties, special talent. Admitted to one college.	Admission exam cutoff.	Jia & Li (2021); Li, Meng, Shi, & Wu (2012)
Colombia	Decentralized	Public and private colleges w/ most programs (selective and non-selective), public and private technical colleges, sub-BA technical colleges.	Set by colleges, publics half to one-third that of privates.	Apply to college and major.	Mostly national exam score (Saber 11). Sometimes combined with high school GPA and interviews.	Admission exam cutoff, college-specific admissions exam for selective public colleges.	González-Velosa et al. (2015); Saavedra (2009); MacLeod et al. (2017); Barrera-Osorio & Bayona-Rodríguez (2019)
France	Decentralized	Grandes ecoles (most selective), vocational institutes, vocational degrees in lycées, universities; mostly public; specialized and multidisciplinary.	Very low.	Apply to college and major. Centralized application system for some colleges (separate admissions decisions).	Must pass high school exit exam. Local priority.	High school exit exam cutoff.	Canaan & Mouganie (2018)
India	Decentralized	Public (central, state, deemed, and institutes of national importance) and private colleges; public colleges are more prestigious.	Public colleges heavily subsidized, not for private. Fees depend on major.	Apply to colleges simultaneously, must be above unknown admission cutoff score. Students choose.	National exam score (there are a few) for certain majors and colleges, sometimes combined with GPA or interview.	Public college admission exam cutoff, affirmative action quotas	Sekhri (2020); Bagde et al. (2016); Bertrand et al. (2010)
Italy	Decentralized	Mostly public, few private and few vocational.	Very low or free.	Apply to college and major.	College-specific admission exam (for some colleges).	Admission exam cutoff.	Anelli (2020)
Kenya	Centralized	Public, private, and vocational colleges	Relatively expensive	Centralized admissions to public universities	Minimum secondary school grades.		Kimenyi et al. (2006)
Nigeria	Centralized	Public (federal and state) and private universities (mostly small and faith-based), Polytechnics, Monotechnics, and Colleges of education.	Free at public (excluding fees)	Apply to up to six institutions: two universities, two polytechnics, and two colleges of education, with first and second choice programs for each	Most universities require minimum test scores for admission.		Oyelere (2011); Aromolaran (2005)
Norway	Centralized	Public (universities, university colleges), some private colleges	Free at public.	Rank up to 15 college-major in single application.	Unknown high school GPA.	High school GPA cutoff. Admitted to one college-major.	Kirkeboen, Leuven, and Mogstad (2016)
South Korea	Decentralized	Universities (national, specialized) and colleges (technical and associates), mostly private but some public.	Expensive	Early and regular admissions, apply separately to colleges. Limit to # of applications in college category.	National exam score (mostly for reguar, not early decision) and high school GPA; quotas	National and regional quotas at each college.	Kim (2021)
UK	Decentralized	Mostly public/state-financed, some private colleges.	Expensive (used to be free), loans readily available.	Centralized application process. List up to 5 college-majors (not ranked). Don't yet know exam scores. After accepted, choose top and insurance college.	Required high school exit exams (e.g., A-levels). Some college-specific exams.	Dale and Krueger (2002) approach.	Broecke (2012); Bratti (2002)

Table 4. Individual Fixed Effects Estimates of the Return to an Associate Degree

State	Quarterly Earnings 5-9 Years After College Entry	
	Men	Women
Arkansas	\$290	\$1,040
California	\$1,650	\$1,650
Kentucky	\$1,740	\$2,270
Michigan	\$1,560	\$2,540
North Carolina	\$1,260	\$2,120
Ohio	\$1,420	\$1,810
Virginia	\$910	\$1,350
Washington	\$480	\$1,000

Source: Belfield and Bailey (2017). Estimates for California are pooled by gender. All estimates are in 2014 dollars.

Table 5. Summary of U.S. Research on the Returns to College Quality

Paper	Method	Data	Colleges Considered	Main Estimate(s)
Andrews, Li, and Lovenheim (2016)	Controls for observable ability and background characteristics	Texas administrative data	Public universities and community colleges in Texas	Substantial heterogeneity in the returns to college quality. Across the income distribution, returns increase for UT-Austin graduates but decrease for Texas A&M graduates. Returns for community college graduates are negative relative to non-flagship 4-year university graduates.
Behrman, Rosenzweig, and Taubman (1996)	Comparisons between twin pairs	Survey of female twin pairs born in Minnesota between 1936-1955	Four-year colleges and universities	PhD-granting, private universities, with well-paid faculty, and smaller enrollments produce students with much higher earnings later in life.
Black, Denning, and Rothstein (2020)	Difference-in-differences comparing students gaining and losing access to UT-Austin to students unlikely to be admitted to a Texas flagship before or after the Texas Top Ten Percent rule.	Texas administrative data	UT-Austin	Students who gain access to UT-Austin experience increases in college enrollment and graduation, with suggestive evidence of higher earnings. Students who lose access to UT-Austin attend less selective colleges but do not see decreases in college enrollment, graduation, or earnings.
Black and Smith (2004)	Propensity score matching	NLSY-79	Four-year colleges and universities across the US	Results suggest significant returns to college quality. Propensity score matching estimates are similar to OLS estimates for men, but smaller for women.
Black and Smith (2006)	Controls for observable ability and background characteristics, and uses an index for college quality and accounts for measurement error	NLSY-79	Four-year colleges and universities across the US	Results suggest significant returns to measures of college quality.
Bleemer (2021)	Regression discontinuity on GPA	California administrative data	Selective institutions in the University of California system	More selective university enrollment caused increases in degree attainment by 30 percentage points and annual early-career earnings by up to \$25,000.
Brewer, Eide, and Ehrenberg (1999)	Choice model of college type controlling for individual characteristics	NLS-1972 and High School and Beyond 1982 cohort	Elite private and public institutions, and middle-rated private institutions	Significant economic return to elite private institutions; smaller premium to attending a middle-rated private institution relative to lower-quality public colleges. Less evidence of a return to attending an elite public institution.
Cohodes and Goodman (2014)	Regression discontinuity on test scores for merit aid eligibility at in-state public colleges in Massachusetts	Massachusetts administrative data	In-state public colleges in Massachusetts	Students induced to attend in-state public colleges in Massachusetts enrolled in lower quality colleges and had lower graduation rates.
Dale and Krueger (2002)	Controls for application portfolio	College and Beyond survey and NLS-72	30 highly selective US colleges	No overall effect of college quality on earnings, but large effect for low-income students.
Ge, Isaac, and Miller (2018)	Controls for application portfolio, but extends Dale and Krueger (2002) by examining new outcomes and not restricting to full-time fully-year workers	College and Beyond survey	30 highly selective US colleges	Results for men echo Dale and Krueger (2002). Results for women are quite different: attending a college with a 100-point higher average SAT score increases women's probability of advanced degree attainment by 5 percentage points and earnings by 14 percent, while reducing their likelihood of marriage by 4 percentage points.
Goodman, Hurwitz, and Smith (2017); Smith, Goodman, and Hurwitz (2020)	Regression discontinuity on SAT score	All SAT takers in Georgia, 2004-2008, National Student Clearinghouse, Equifax credit bureau.	All public universities in Georgia.	Low SAT scoring disadvantaged students who enroll in a public university are 41 pp more likely to earn BA. Household income at about age 30 increases by 20%.
Hoekstra (2009)	Regression discontinuity on SAT score	Administrative data from a large flagship state university	One large flagship state university (the most selective in the state)	Attending the most selective college in this state increased earnings 10-15 years after high school graduation by 20% for white men.
Kozakowski (2020)	Regression discontinuity on SAT score and GPA	All public school students in Massachusetts	Massachusetts public colleges	Gaining admission increases BA completion by 15 pp and earnings by 26%.
Long (2008)	Compares results across three methods: 1) distance instrumental variable, 2) controlling for application portfolio, and 3) matching.	National Education Longitudinal Survey	Four-year colleges and universities across the US	Each methodology finds that college quality has positive significant effects on degree attainment and labor market earnings.
Mountjoy and Hickman (2021)	Dale and Krueger controls for application portfolio	Texas administrative data	Texas public universities	Small effects of college quality (relative to the literature and raw means): a one standard deviation increase in the college value-added distribution corresponds to a 3.7 percentage point increase in bachelor's degree completion and a \$1,300 (3%) increase in annual earnings at age 28.
Reynolds (2012)	Propensity score matching comparing students who begin their college career at a two-year college instead of a four-year college.	National Education Longitudinal Study of 1988	Two-year and four-year institutions in the US	Estimates show large negative impacts on educational attainment and labor market outcomes for men and women who begin at a two-year college.
Smith (2013)	Comparisons between twin pairs, with Dale and Krueger portfolio controls.	SAT takers and National Student Clearinghouse	Four-year colleges and universities across the US	Attending a college with 100 points higher median SAT score increases bachelor's degree completion within 4 years by 5 percentage points.
Zimmerman (2014)	Regression discontinuity on high school GPA	Administrative education data from 15 Florida counties, Florida University System, state UI	Single Florida public college.	22% earnings gain between 8-14 years after high school, mostly for males and low-income students.

Table 6. Summary of Research on the Returns to For-Profit College Enrollment

Paper	Level	Data	Methods	Comparison	Results
Deming, Goldin, & Katz (2012)	2-year & 4-year	BPS	Selection on observables	FP enrollment relative to other college students	FP students have lower earnings and higher likelihood of unemployment
Lang & Weinstein (2013)	2-year	BPS	Completers vs. non-completers across institution type	FP completion effect relative to other college completion effect	No differential returns to completing an FP degree or credential
Cellini & Chaudhary (2014)	2-year	NLSY-97	Individual fixed effects	For-profit enrollment vs. non-college attendees	For-profit students have 10% higher earnings than high school graduates (conditional on employment)
Denice (2015)	2-year & 4-year	NLSY-97	Selection on observables	For-profit completion vs. other college completion & non-attendees	2-year for-profit completion leads to lower earnings relative to other 2-year completion and no difference from non-attendees earnings. 4-year for-profit completers have similar earnings to other college completers.
Liu and Belfield (2014a)	2-year & 4-year	UI & NSC	Individual fixed effects among transfer students	For-profit vs. public enrollees	Students who transfer from public to for-profit experience lower earnings than those who transfer to other colleges
Liu and Belfield (2014b)	2-year & 4-year	ELS	Selection on observables	For-profit vs. public enrollees	For-profit enrollees have lower earnings than public college enrollees

Hoxby (2015)	4-year	IRS & College Board Data	Selection on observables (matched comparisons)	For-profit vs. other 4-year students	For-profit enrollees have lower earnings than other college enrollees
Cellini & Turner (2019)	2-year	IRS & Dept. of Education Data	Matched difference-in-difference: effect of for-profit vs. other college enrollment.	For-profit certificate students relative to other public college certificate students	For-profit credentials lead to lower earnings. Lower returns for online for-profit students.
Jepsen, Mueser, & Jeon (2016)	2-year	Administrative education and UI data from Missouri	Individual fixed effects	Change in earnings surrounding for-profit completion	For-profit completion leads to an increase in earnings
Armona, Chakrabarti, & Lovenheim (Forthcoming)	2-year & 4-year	IPEDS, Dept. of Education data, College Scorecard	IV: labor demand shocks interacted with for-profit college supply	For-profit enrollees relative to public enrollees	Negative earnings effect of for-profit (not significant). For-profit students take out more loans and default at higher rates
Darolia et al. (2015)	2-year	Author generated data	Randomized audit study	For-profit vs. other 2-year credentials	No significant difference in job callback rates
Deming et al. (2016)	2-year & 4-year	Author generated data	Randomized audit study	For-profit vs. other 2-year & 4-year credentials/degrees	No significant difference in job callback rates for AA degrees. Negative effect of for-profit 4-year degrees and 2-year certificates relative to public

Source: Adapted from Cellini and Koedel (2017). BPS=Beginning Postsecondary Study; ELS=Educational Longitudinal Study; NLSY-97=National Longitudinal Survey of Youth, 1997; UI=unemployment insurance system quarterly earnings data; NSC=national student clearinghouse.

Table 7. Summary of International Research on Returns to College Quality

Paper	Country	Method	Setting	Main Result
Francis & Tannuri-Pianto (2012)	Brazil	Quotas	Race-based affirmative action to selective college.	Increase in Black student enrollment to selective college.
Hastings, Neilson, & Zimmerman (2013)	Chile	RD	All Chile colleges and majors. Observe counterfactual choice.	Enroll --> 4.5% - earnings, increasing in selectivity.
Rodríguez, Urzúa, & Reyes (2016)	Chile	Select on Obs	Universe of 2008 higher ed graduates, earnings in 2011.	5-year degree has positive avg. earnings, ~35% negative net returns.
Chen (2019)	China	Resume Audit	Fictitious resumes to Chinese firms, varying college selectivity.	More selective Chinese (US) colleges --> 13% (9%) more callbacks.
Jia & Li (2021)	China	RD	Admission to elite college sector (~ top 100 colleges).	Enroll --> 28-45% - earnings.
Li, Meng, Shi, & Wu (2012)	China	Select on Obs	Elite college sector (~ top 100 colleges).	Enroll --> 10.7% - earnings.
Saavedra (2009)	Columbia	RD	Admission to top-ranked college.	Enroll --> 25% - earnings, 0.2 SD -exit exam scores.
MacLeod et al. (2017)	Columbia	Select on Obs, Economic Theory	Introduction of college exit exam and first job.	Positive return to quality and wage growth faster in high quality colleges.
Canaan & Mouganie (2018)	France	RD	Low-skilled students, passing HS exit exam.	0.13 SD - peer scores, 12.5% - earnings.
Anelli (2020)	Italy	RD	Admission to selective private university.	Enroll --> 79% - earnings.
Kirkeboen, Leuven, and Mogstad (2016)	Norway	RD	All Norway colleges and majors. Observe counterfactual choice.	College large premium for one selective college but not others, most wage premium from field of study.
Kim (2021)	South Korea	Quotas	National and regional quotas, along with cohort swells.	4-year college enroll --> 63-75% - hourly wages.
Lindahl & Regnér (2005)	Sweden	Siblings	Enrollment to free colleges for 19,000 siblings.	Variation in wage premium by college, highest for oldest colleges.
Broecke (2012)	UK	Dale and Kruege	Nearly all UK colleges.	1 SD - selectivity, 7% - earnings.

Table 8. Individual Fixed Effects Studies on the Return to Sub-Baccalaureate Programs

Panel A: Men				
Study	State(s)	AA	Diploma	Certificate
Jepsen, Troske, and Coomes (2014)	Kentucky	\$1,484 (24.6%)	\$1,265 (20.6%)	\$297 (4.8%)
Stevens, Kurlaender, and Grosz (2019)	California	37.4%	20.1%	13.3%
Bahr et al. (2015)	Michigan	\$1,441 (11.8%)	\$918 (3.6%)	\$1,345 (8.4%)
Dadger and Trimble (2015)	Washington State	6.3%	14.9%	-2.8%
Bettinger and Soliz (2016)	Ohio	21.2%	18.0%	41.1%
Xu and Trimble (2016)	North Carolina	\$1,256	\$953	\$278
Xu and Trimble (2016)	Virginia	\$773	\$200	\$153
Carruthers and Sanford (2018)	Tennessee		\$1,034 (17.3%)	\$292 (1.2%)
Panel B: Women				
Study	State(s)	AA	Diploma	Certificate
Jepsen, Troske, and Coomes (2014)	Kentucky	\$2,363 (55.7%)	\$1,914 (45.1%)	\$299 (7.0%)
Stevens, Kurlaender, and Grosz (2019)	California	34.3%	26.3%	9.6%
Bahr et al. (2015)	Michigan	\$2,346 (32.7%)	\$620 (13.2%)	\$268 (0.4%)
Dadger and Trimble (2015)	Washington State	20.8%	1.3%	-0.3%
Bettinger and Soliz (2016)	Ohio	26.4%	21.7%	5.5%
Xu and Trimble (2016)	North Carolina	\$1,256	\$953	\$278
Xu and Trimble (2016)	Virginia	\$773	\$200	\$153
Carruthers and Sanford (2018)	Tennessee		\$1,034 (17.3%)	\$292 (1.2%)

Notes: Estimates for Xu and Trimble (2016) are pooled by gender and are listed in both panels. When only percents are listed the estimates in the paper are from log earnings regressions. When levels are shown, the paper presents estimates in dollars and percent estimates are calculated (shown in parentheses) using mean earnings when available.

Table 9. Individual Fixed Effects Studies on the Return to Sub-baccalaureate Programs by Field of Study

Panel A: Men				
Study	State	AA	Diploma	Certificate
Jepsen, Troske, and Coomes (2014)	Kentucky	Health: 0.715 Business: -0.027	Health: 0.412 Business: -0.217	Health: 0.006 Business: -0.002
Stevens, Kurlaender, and Grosz (2019)	California	Health: 0.669 Business: 0.156 IT: 0.140 PS: 0.148	Health: 0.314 Business: 0.006 IT: -0.037 PS: 0.144	Health: 0.118 Business: 0.179 IT: 0.099 PS: 0.161
Bahr et al. (2015)	Michigan	Health: 0.597 Business: -0.053 Nurse: 0.862 IT: 0.177 PS: 0.146	Health: Business: -0.042 Nurse: IT: 0.067 PS:	Health: 0.006 Business: Nurse: IT: 0.288 PS:
Dadgar and Trimble (2015)	Washington State	Health: 0.139 Business: 0.011 Nurse: 0.272 IT: -0.010 PS: 0.090	Health: -0.016 Business: -0.142 Nurse: 0.199 IT: -0.029 PS: 0.003	Health: 0.010 Business: 0.041 Nurse: -0.093 IT: -0.056 PS: 0.221
Bettinger and Soliz (2016)	Ohio	Health: 0.503 Business: 0.129 IT: 0.164 PS: 0.139	Health: 0.220 Business: 0.176 IT: PS:	Health: 0.178 Business: 0.051 IT: 0.253 PS: 0.653
Panel B: Women				
Study	State	AA	Diploma	Certificate
Jepsen, Troske, and Coomes (2014)	Kentucky	Health: 1.209 Business: 0.179	Health: 0.669 Business: 0.043	Health: 0.103 Business: 0.047
Stevens, Kurlaender, and Grosz (2019)	California	Health: 0.696 Business: 0.131 IT: 0.052 PS: 0.167	Health: 0.447 Business: 0.141 IT: 0.255 PS: 0.183	Health: 0.110 Business: 0.166 IT: 0.078 PS: 0.044
Bahr et al. (2015)	Michigan	Health: 0.657 Business: 0.035 Nurse: 1.030 IT: 0.060 PS: 0.017	Health: 0.057 Business: 0.192 Nurse: IT: PS:	Health: 0.019 Business: -0.239 Nurse: 0.019 IT: PS:
Dadgar and Trimble (2015)	Washington State	Health: 0.144 Business: 0.045 Nurse: 0.377 IT: 0.041 PS: 0.152	Health: 0.062 Business: 0.024 Nurse: 0.293 IT: 0.037 PS:	Health: -0.034 Business: 0.075 Nurse: -0.056 IT: -0.047 PS: -0.000
Bettinger and Soliz (2016)	Ohio	Health: 0.503 Business: 0.129 IT: 0.164	Health: 0.220 Business: 0.176 IT:	Health: 0.178 Business: 0.051 IT: 0.253

		PS: 0.139	PS:	PS: 0.653
--	--	-----------	-----	-----------

Notes: PS = Protective Services; IT = Information Technology. All estimates reflect percentage changes in earnings from a given credential; level estimates are converted into percent effects using gender-specific means. For Stevens, Kurlaender, and Grosz (2019), the Diploma estimates are programs that require 30-59 Units and the Certificate estimates are programs that require 6-17 units. The Health estimates in Jepsen, Troske, and Coomes (2014), Stevens, Kurlaender, and Grosz (2019), and Bettinger and Soliz (2016) include Nursing. Estimates in Bettinger and Soliz (2016) are pooled by gender.

Table 10. Highest and Lowest Paying Majors, Prime-Age Earners with a BA from the 2014-2019 ACS

10 Highest Paying Fields		10 Lowest Paying Fields	
Field of Study	Average Earnings	Field of Study	Average Earnings
Electrical Engineering	\$106,093	Early Childhood Education	\$39,542
Engineering Mechanics		Cosmetology Services And	
Physics And Science	\$106,252	Culinary Arts	\$43,060
Chemical Engineering	\$107,042	Miscellaneous Fine Arts	\$44,744
Aerospace Engineering	\$107,268	Family And Consumer Sciences	\$44,892
Nuclear Engineering	\$107,923	Studio Arts	\$45,196
Economics	\$108,799	Human Services And Community	
		Organization	\$45,494
Zoology	\$111,240	Social Work	\$45,921
Actuarial Science	\$112,827	Visual And Performing Arts	\$46,376
Health And Medical		Elementary Education	
Preparatory Programs	\$128,713	Theology And Religious	\$46,453
Petroleum Engineering	\$131,063	Vocations	\$47,297

Source: 2014-2019 American Community Survey. Estimates are means among prime-age workers (age 25-64) with a BA.

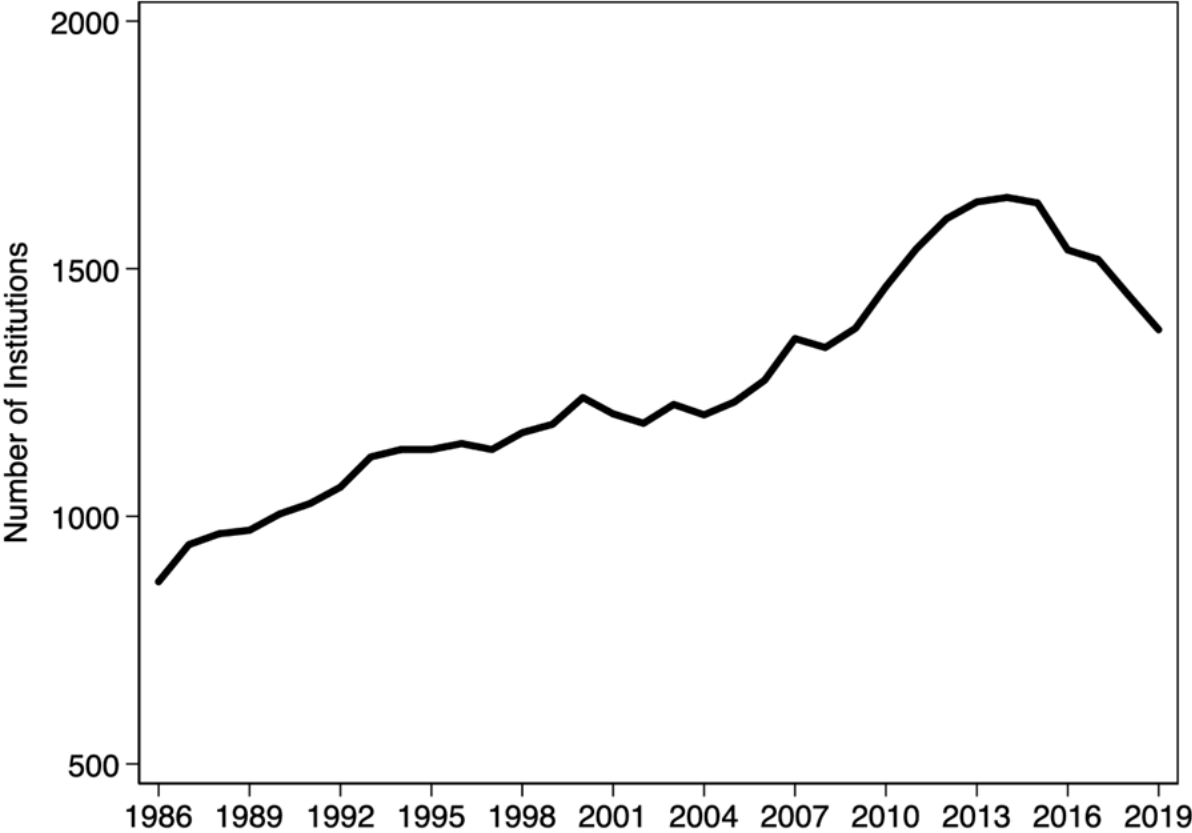
Table 11. Estimates of the Return to Three Graduate Degrees: MBA, MD, and JD

Panel A: Business (MBA)			
Study	Data/Sample	Method	Return
Altonji and Zhong (2021)	National Survey of College Graduates; National Survey of Recent College Graduates	FE-cg	0.162
Altonji and Zhu (2021)	Linked K-12, Higher Education, and UI Data in Texas	FE-cg	0.162
Altonji and Zhu (2021)	Linked K-12, Higher Education, and UI Data in Texas	Individual FE	0.210
Arcidiacono, Cooley, and Hussey (2008)	Survey of registrants for the Graduate Management Admissions Test (GMAT)	Individual FE	0.0484 (men) 0.1956 (men; top 10 program) 0.0378 (women) 0.1678 (women; top 10 program)
Grove and Hussey (2011)	Survey of registrants for the Graduate Management Admissions Test (GMAT)	Individual FE	0.0726 0.1305 (MBA-finance) 0.1504 (MBA-MIS)
Grove and Hussey (2014)	Survey of registrants for the Graduate Management Admissions Test (GMAT)	Individual FE	0.092 for each standard deviation increase in law school quality index.
Panel B: Law (JD)			
Study	Data/Sample	Method	Return
Altonji and Zhong (2021)	National Survey of College Graduates; National Survey of Recent College Graduates	FE-cg	0.462
Altonji and Zhu (2021)	Linked K-12, Higher Education, and UI Data in Texas	FE-cg	0.568
Altonji and Zhu (2021)	Linked K-12, Higher Education, and UI Data in Texas	Individual FE	0.464
Simkovic and McIntyre (2014)	Survey of Income and Program Participation (SIPP)	Control for college major and HS curriculum	0.61 0.57 (men) 0.70 (women)
Oyer and Schaefer (2019)	After the JD (AJD); those who pass the bar in 2000.	OLS, controlling for undergrad GPA and institution	0.120 (top 10 program) -0.077 (21-100 program) Both relative to 11-20 ranked program
Sander and	After the JD (AJD).	OLS,	0.27 (top 10 program)

Bambauer (2012)		controlling for demographics and broad undergrad major	0.16 (11-20 program) Negative for 51+ ranked programs All relative to 36-50 ranked programs
Panel C: Medicine (MD)			
Study	Data/Sample	Method	Return
Altonji and Zhong (2021)	National Survey of College Graduates; National Survey of Recent College Graduates	FE-cg	0.619
Altonji and Zhu (2021)	Linked K-12, Higher Education, and UI Data in Texas	FE-cg	0.761
Altonji and Zhu (2021)	Linked K-12, Higher Education, and UI Data in Texas	Individual FE	0.738
Ketel, Leuven, Oosterbeek, and Van Der Klauww, (2016)	Administrative data from the Netherlands	IV using admission lotteries to medical programs	0.20 per year 0.48 after 22 years
Ketel, Leuven, Oosterbeek, and Van Der Klauww, (2019)	Administrative data from the Netherlands	IV using admission lotteries to dental programs	0.49 after 22 years

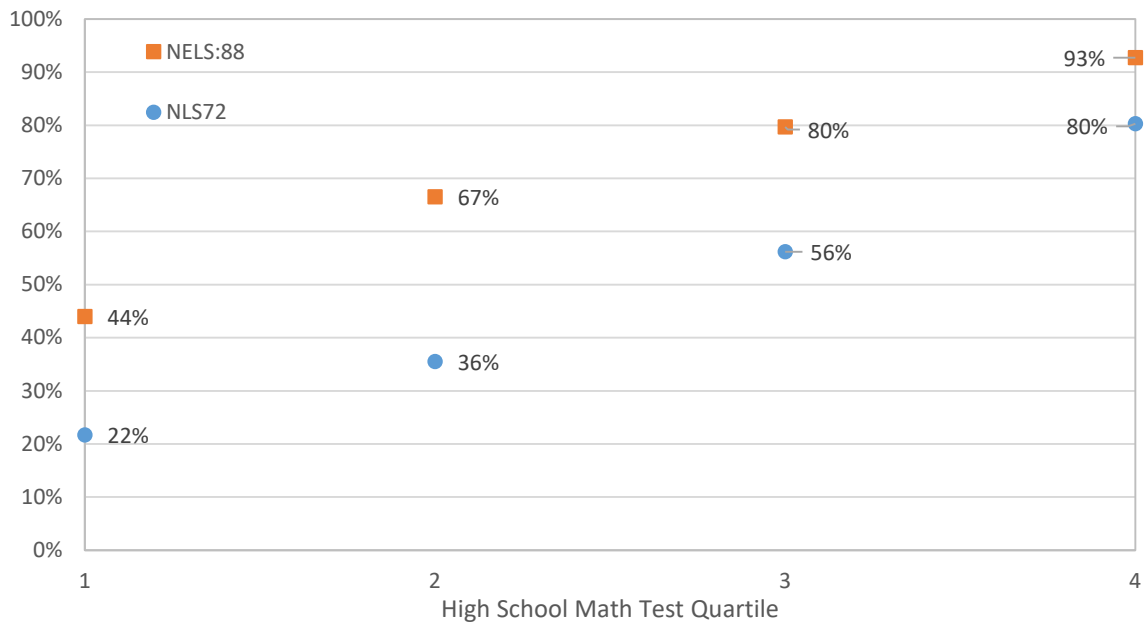
Note: MIS = management information systems. All program ranks refer to rankings used in the individual studies.

Figure 1: The Number of Postsecondary Institutions Awarding Associates and Bachelors Degrees over Time

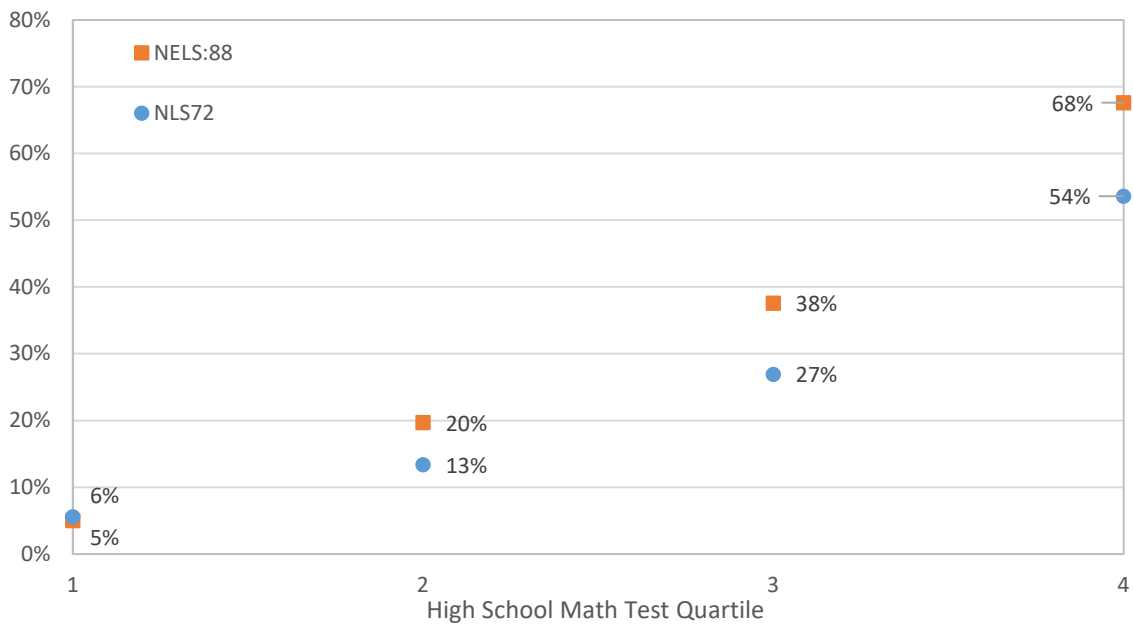


Source: Authors calculations from IPEDS. Includes any institution that awards both associates and bachelor's degrees.

Figure 2. BA Completion Rates by Math Test Quartile, NELS:88
Panel A. College Enrollment Conditional on HS Completion

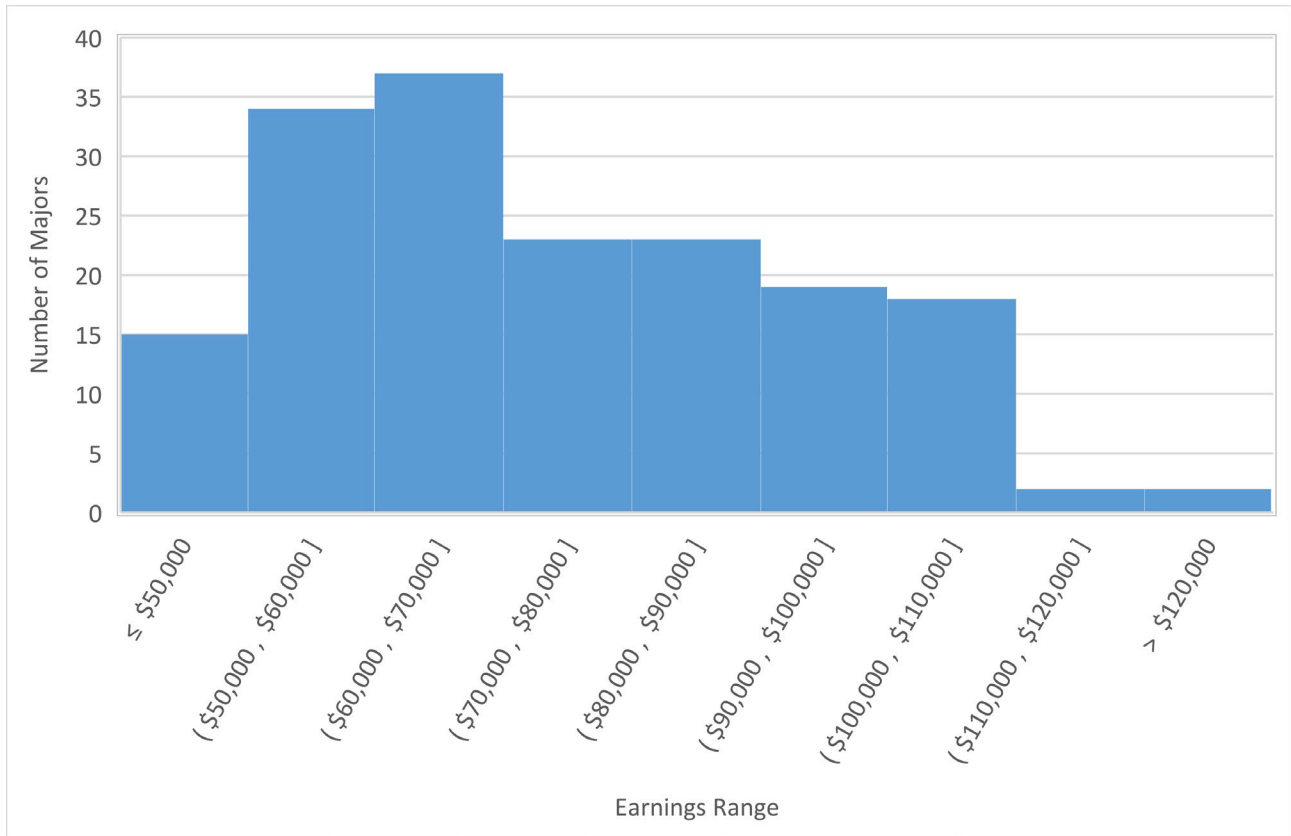


Panel B. BA Completion Conditional on HS Completion



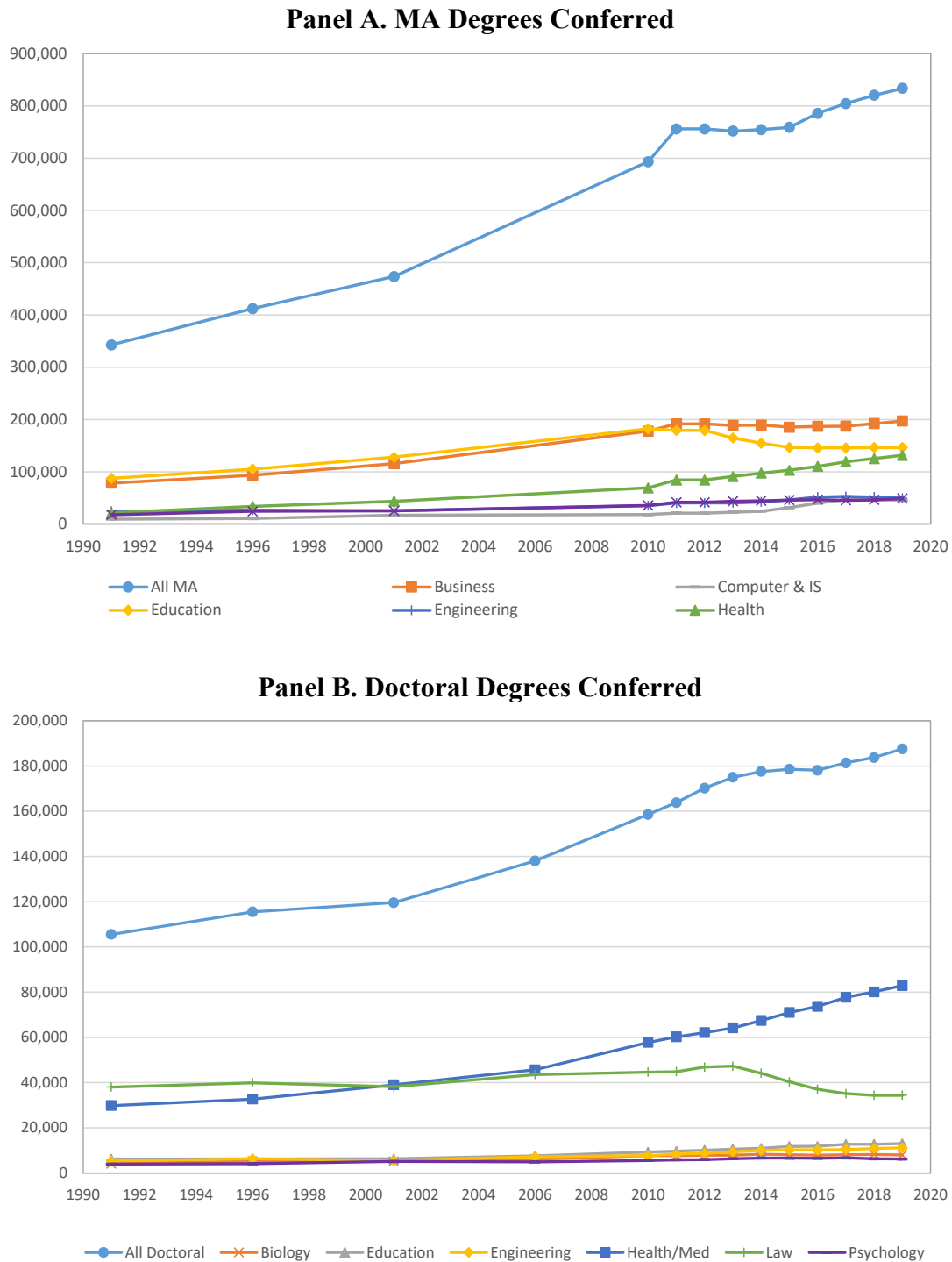
Source: Adapted from Bound, Lovenheim, and Turner (2010), Figure 2. Estimates are from the National Longitudinal Study of the High School Class of 1972 (NLS72) and the National Educational Longitudinal Study of 1988 (NELS:88).

Figure 3. Distribution of Mean Earnings by College Major,



Source: 2014-2019 American Community Survey among prime-age workers (age 25-64) with BA degrees.

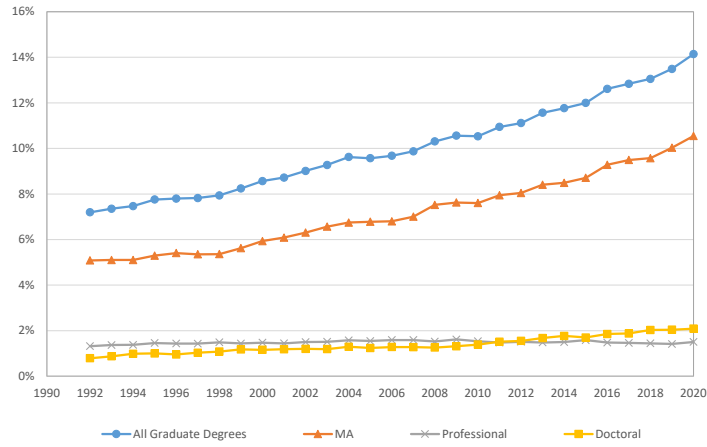
Figure 4. MA and Doctoral Degrees Conferred in the US, Overall and by Selected Field



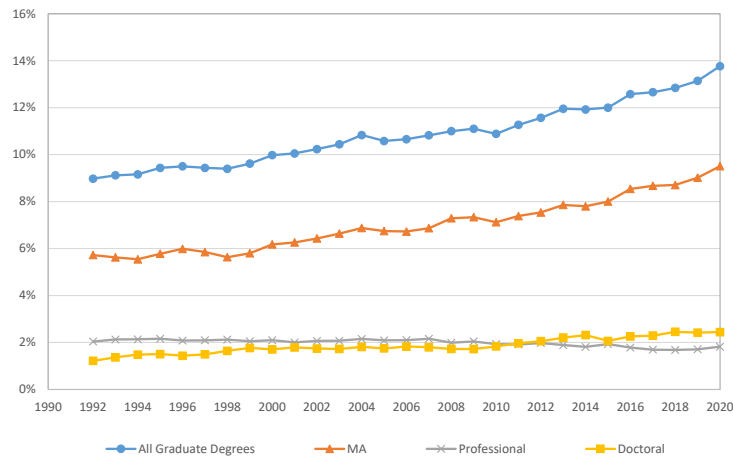
Source: Digest of Education Statistics Tables 323.10 (Panel A) and 324.10 (Panel B). All years refer to the spring of the given academic year. The six largest fields of study are included in each figure for each degree level. Health/Med doctoral degrees include MD and DDS degrees, and Law doctoral degrees include JD degrees.

Figure 5. Percent of the 25+ US Population with a Graduate Degree, Overall and by Gender

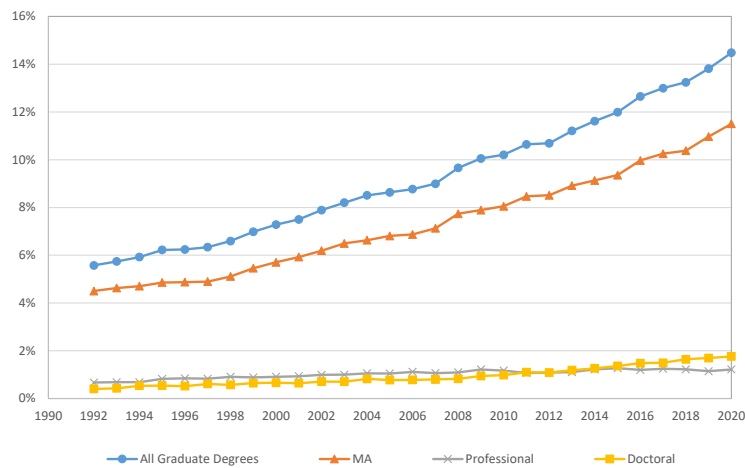
Panel A. Men and Women



Panel B. Men



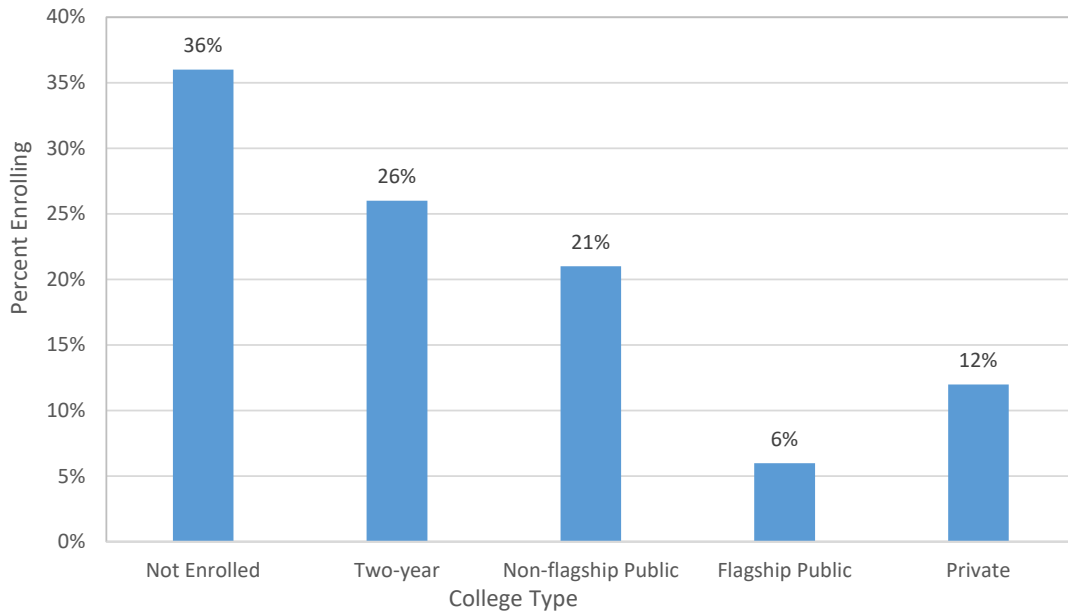
Panel C. Women



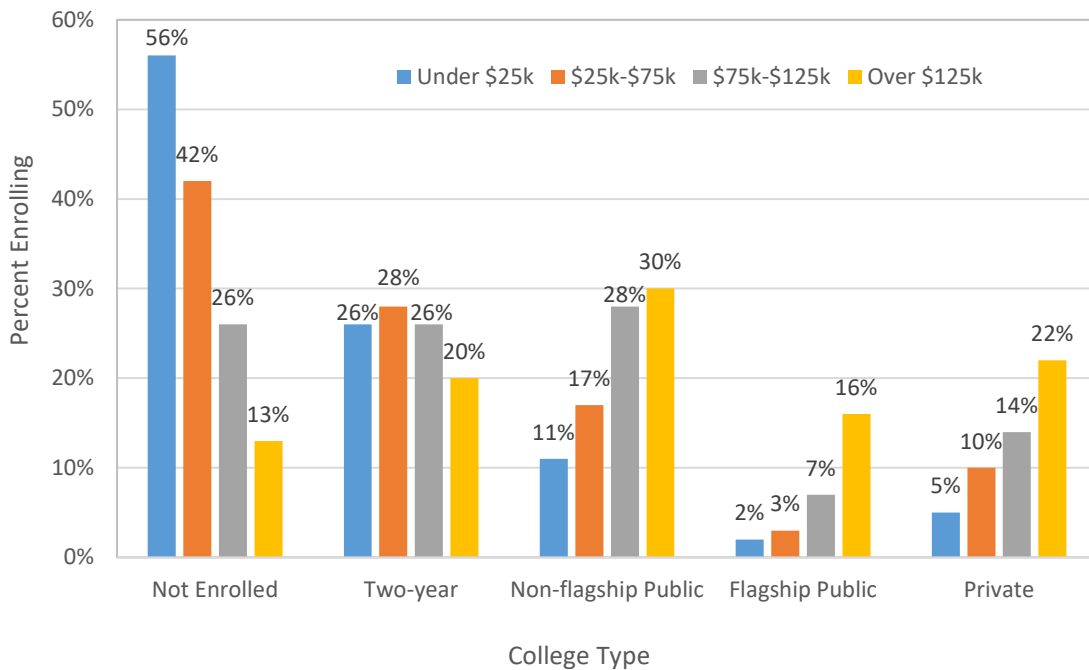
Source: Current Population Survey data for US residents aged 25+.

Figure 6. College Enrollment within Two Years of High School Graduation by Family Income, NLSY97

Panel A. All High School Graduates



Panel B. By High Parental Income in 1997



Source: Adapted from Lovenheim (2011) using data from the National Longitudinal Survey of Youth 1997.

Works Cited

- Altonji, J. G., Arcidiacono, P., & Maurel, A. (2016). The analysis of field choice in college and graduate school: Determinants and wage effects. In *Handbook of the Economics of Education* (Vol. 5, pp. 305-396). Elsevier.
- Altonji, J. G., Blom, E. & Meghir, C. (2012). Heterogeneity in human capital investments: high school curriculum, college major, and careers. *Annu. Rev. Econ.*, 4(1), 185-223.
- Altonji, J. G., Kahn, L.B. & Speer, J.D. (2016). Cashier or Consultant? Entry labor market conditions, field of study, and career success. *Journal of Labor Economics*, 34(S1), S361-S401.
- Altonji, J. G. & Zhong, L. (2021). The labor market returns to advanced degrees. *Journal of Labor Economics*, 39(2), 303-360.
- Altonji, J. G. & Zhu, Z. (2021). Returns to specific graduate degrees: Estimates using Texas administrative records. Unpublished working paper.
- Andrews, R. J., Imberman, S. A. & Lovenheim, M. F. (2017). Risky business? The effect of majoring in business on earnings and educational attainment. NBER working paper, No. w23575.
- Andrews, R. J., Imberman, S. A. & Lovenheim, M. F. (2020). Recruiting and supporting low-income, high-achieving students at flagship universities. *Economics of Education Review*, 74: 101923.
- Andrews, R. J., Li, J., & Lovenheim, M. F. (2014). Heterogeneous paths through college: Detailed patterns and relationships with graduation and earnings. *Economics of Education Review*, 42, 93-108.
- Andrews, R. J., Li, J., & Lovenheim, M. F. (2016). Quantile treatment effects of college quality on earnings. *Journal of Human Resources*, 51(1), 200-238.
- Andrews, R. J., & Stange, K. M. (2019). Price regulation, price discrimination, and equality of opportunity in higher education: Evidence from Texas. *American Economic Journal: Economic Policy*, 11.4, 31-65.
- Anelli, M. (2020). The returns to elite university education: A quasi-experimental analysis. *Journal of the European Economic Association*, 18(6), 2824-2868.
- Angrist, J., Autor, D., Hudson, S., & Pallais, A. (2017). Evaluating post-secondary aid: Enrollment, persistence, and projected completion effects. NBER working paper, No. w23015.
- Angrist, J., Lang, D., & Oreopoulos, P. (2009). Incentives and services for college achievement: Evidence from a randomized trial. *American Economic Journal: Applied Economics*, 1(1), 136-63.
- Arcidiacono, P. (2004). Ability sorting and the returns to college major. *Journal of Econometrics*, 121(1-2), 343-375.
- Arcidiacono, P., & Lovenheim, M. (2016). Affirmative action and the quality-fit trade-off. *Journal of Economic Literature*, 54(1), 3-51.
- Arcidiacono, P., Aucejo, E., Maurel, A. & Ransom, T. (2016). College attrition and the dynamics of information revelation. NBER working paper, No. 22325.

- Arcidiacono, P., Aucejo, E. M. & Hotz, V. J. (2016). University differences in the graduation of minorities in STEM fields: Evidence from California. *American Economic Review*, 106(3), 525-62.
- Arcidiacono, P., Cooley, J. & Hussey, A. (2008). The economic returns to an MBA. *International Economic Review*, 49(3), 873-899.
- Armona, L., Chakrabarti, R., & Lovenheim, M. F. (Forthcoming). How does for-profit college attendance affect student loans, defaults and earnings? *Journal of Financial Economics*.
- Ashenfelter, O. (1978). Estimating the effect of training programs on earnings. *Review of Economics and Statistics*, 47-57.
- Ashenfelter, O., & Krueger, A. (1994). Estimates of the economic return to schooling from a new sample of twins. *American Economic Review*, 84(5), 1157-1173.
- Autor, D. H. (2014). Skills, education, and the rise of earnings inequality among the “other 99 percent”. *Science*, 344(6186), 843-851.
- Autor, D. H., Katz, L. F., & Kearney, M. S. (2008). Trends in U.S. wage inequality: Revising the revisionists. *Review of Economics and Statistics*, 90(2), 300–323.
- Autor, D. H., Levy, F. & Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. *Quarterly Journal of Economics*, 118(4), 1279-1333.
- Avery, C., Fairbanks, A., & Zeckhauser, R. J. (2009). *The early admissions game*. Harvard University Press.
- Avery, C., & Levin, J. (2010). Early admissions at selective colleges. *American Economic Review*, 100(5), 2125-56.
- Bahr, P. R., Dynarski, S., Jacob, B., Kreisman, D., Sosa, A., & Wiederspan, M. (2015). Labor market returns to community college awards: Evidence from Michigan. A CAPSEE working paper. Center for Analysis of Postsecondary Education and Employment.
- Barr, A. & Castleman, B. (2018). An engine of economic opportunity: Intensive advising, college success, and social mobility. *Texas A&M Working Paper*.
- Barr, A., Bird, K.A., Castleman, B.L., & Skimmyhorn, W.L. (2021). Can information and advising affect postsecondary participation and attainment for non-traditional students? Evidence from a large-scale experiment with the U.S. Army. *EdWorkingPaper No. 22-529*.
- Barr, A. & Castleman, B. (2021). The bottom line on college advising: Large increases in degree attainment. *EdWorkingPaper No. 21-481*.
- Barr, A., Kawano, L., Sacerdote, B., Skimmyhorn, W., & Stevens, M. (2021). You can't handle the truth: The effects of the post-9/11 GI bill on higher education and earnings. NBER working paper, No. w29024
- Bartik, T. J., Hershbein, B., & Lachowska, M. (2021). The effects of the Kalamazoo Promise Scholarship on college enrollment and completion. *Journal of Human Resources*, 56(1), 269-310.
- Becker, G. S. (1962). Investment in human capital: A theoretical analysis. *Journal of Political Economy*, 70(5), Part 2: 9-49.
- Behrman, J. R., Rosenzweig, M. R., & Taubman, P. (1996). College choice and wages: Estimates using data on female twins. *The Review of Economics and Statistics*, 672-685.

- Belfield, C., & Bailey, T. (2017). The Labor Market Returns to Sub-Baccalaureate College: A Review. A CAPSEE Working Paper. *Center for Analysis of Postsecondary Education and Employment*.
- Bertrand, M., Goldin, C. & Katz, L. F. (2010). Dynamics of the gender gap for young professionals in the financial and corporate sectors. *American Economic Journal: Applied Economics*, 2(3), 228-55.
- Bettinger, E. P. & Baker, R. B. (2014). The effects of student coaching: An evaluation of a randomized experiment in student advising. *Educational Evaluation and Policy Analysis*, 36(1), 3-19.
- Bettinger, E. P. & Evans, B. J. (2019). College guidance for all: A randomized experiment in pre-college advising. *Journal of Policy Analysis and Management*, 38(3), 579-599.
- Bettinger, E., Gurantz, O., Kawano, L., Sacerdote, B., & Stevens, M. (2019). The long-run impacts of financial aid: Evidence from California's Cal Grant." *American Economic Journal: Economic Policy*, 11(1), 64-94.
- Bettinger, E. P. & Long, B. T. (2009). Addressing the needs of underprepared students in higher education: Does college remediation work? *Journal of Human Resources*, 44(3), 736-771.
- Bettinger, E. & Soliz, A. (2016). Returns to vocational credentials: Evidence from Ohio's community and technical colleges. A CAPSEE working paper. Center for Analysis of Postsecondary Education and Employment.
- Bhuller, M., Mogstad, M., & Salvanes, K. G. (2017). Life-cycle earnings, education premiums, and internal rates of return. *Journal of Labor Economics*, 35(4), 993-1030.
- Black, D. A., & Smith, J. A. (2004). How robust is the evidence on the effects of college quality? Evidence from matching. *Journal of econometrics*, 121(1-2), 99-124.
- Black, D. A., & Smith, J. A. (2006). Estimating the returns to college quality with multiple proxies for quality. *Journal of labor Economics*, 24(3), 701-728.
- Black, S. E., Cortes, K. E., & Lincove, J. A. (2015). Academic undermatching of high-achieving minority students: Evidence from race-neutral and holistic admissions policies. *American Economic Review*, 105(5), 604-10.
- Black, S. E., Cortes, K. E., & Lincove, J. A. (2020). Apply yourself: Racial and ethnic differences in college application. *Education Finance and Policy*, 15(2), 209-240.
- Black, S. E., Denning, J. T., & Rothstein, J. (Forthcoming). Winners and losers? The effect of gaining and losing access to selective colleges on education and labor market outcomes. *American Economic Journal: Applied Economics*.
- Bleemer, Z. (2021). Top percent policies and the return to postsecondary selectivity. Available at SSRN 3272618.
- Bleemer, Z. & Mehta, A. (Forthcoming). Will studying economics make you rich? A regression discontinuity analysis of the returns to college major. *American Economic Journal: Applied Economics*.
- Bloem, M. D., Pan, W., & Smith, J. (2021). College entrance exam-taking strategies in Georgia. *Southern Economic Journal*, 88(2), 587-627.
- Boatman, A. & Long, B. T. (2018). Does remediation work for all students? How the effects of postsecondary remedial and developmental courses vary by level of academic preparation. *Educational Evaluation and Policy Analysis*, 40(1), 29-58.

- Böckerman, P., Haapanen, M. & Jepsen, C. (2018). More skilled, better paid: Labour-market returns to postsecondary vocational education. *Oxford Economic Papers*, 70(2), 485-508.
- Böckerman, P., Hämäläinen, U. & Uusitalo, R. (2009). Labour market effects of the polytechnic education reform: The Finnish experience. *Economics of Education Review*, 28(6), 672-681.
- Bound, J., Braga, B., Khanna, G., & Turner, S. (2019). Public universities: The supply side of building a skilled workforce. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, 5(5), 43-66.
- Bound, J., Lovenheim, M. F., & Turner, S. (2010). Why have college completion rates declined? An analysis of changing student preparation and collegiate resources. *American Economic Journal: Applied Economics*, 2(3), 129-57.
- Bound, J., & Turner, S. (2011). Dropouts and diplomas: The divergence in collegiate outcomes. In *Handbook of the Economics of Education* (Vol. 4, pp. 573-613). Elsevier.
- Bound, J., Jaeger, D. A., & Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American statistical association*, 90(430), 443-450.
- Bound, J., & Solon, G. (1999). Double trouble: on the value of twins-based estimation of the return to schooling. *Economics of Education Review*, 18(2), 169-182.
- Bowen, W. G., Chingos, M. M., & McPherson, M. S. (2009). *Crossing the finish line*. Princeton University Press.
- Brewer, D. J., Eide, E. R., & Ehrenberg, R. (1999). Does it pay to attend an elite private college? Cross-cohort evidence on the effects of college type on earnings. *Journal of Human Resources*, 34(1), 104-123.
- Broecke, S. (2012). University selectivity and earnings: Evidence from UK data on applications and admissions to university. *Economics of Education Review*, 31(3), 96-107.
- Brunello, G. & Rocco, L. (2017). The effects of vocational education on adult skills, employment and wages: What can we learn from PIAAC?" *SERIEs*, 8(4), 315-343.
- Bucarey, A., Contreras, D. & Muñoz, P. (2020). Labor market returns to student loans for university: Evidence from Chile." *Journal of Labor Economics*, 38(4).
- Bulman, G. (2015). The effect of access to college assessments on enrollment and attainment. *American Economic Journal: Applied Economics*, 7(4), 1-36.
- Calcagno, J. C. & Long, B. T. (2008). The impact of postsecondary remediation using a regression discontinuity approach: Addressing endogenous sorting and noncompliance. NBER working paper, No. w14194.
- Campbell, S., Macmillan, L., Murphy, R., & Wyness, G. (Forthcoming). Matching in the Dark? Inequalities in student to degree match. *Journal of Labor Economics*.
- Canaan, S., & Mouganie, P. (2018). Returns to education quality for low-skilled students: Evidence from a discontinuity. *Journal of Labor Economics*, 36(2), 395-436.
- Card, D. (1995). Using geographic variation in college proximity to estimate the returns to schooling. In L.N. Christofides, et al. (Eds.), *Aspects of Labour Market Behavior: Essays in Honor of John Vanderkamp*, University of Toronto Press, Toronto, 201-221.
- Card, D. (2001). Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica*, 69(5), 1127-1160.

- Carneiro, P. & Heckman, J J. (2002). The evidence on credit constraints in post-secondary schooling. *The Economic Journal*, 112(482), 705-734.
- Carrell, S. & Sacerdote, B. (2017). Why do college-going interventions work? *American Economic Journal: Applied Economics*, 9(3), 124-51.
- Carruthers, C. K., Fox, W. F., & Jepsen, C. (2020). Promise Kept? Free Community College, Attainment, and Earnings in Tennessee. February 2020 (work in progress). *Working paper*.
- Cellini, S. R. (2021). For-profit colleges in the United States: Insights from two decades of research.
- Cellini, S. R., & Chaudhary, L. (2014). The labor market returns to a for-profit college education. *Economics of Education Review*, 43, 125-140.
- Cellini, S. R., & Chaudhary, L. (2020). [Commercials for college? Advertising in higher education](#), The Brookings Institution, Brown Center on Education Policy Report, May 19, 2020.
- Cellini, S. R., & Darolia, R. (2017). High costs, low resources, and missing information: Explaining student borrowing in the for-profit sector. *The ANNALS of the American Academy of Political and Social Science*, 671(1), 92-112.
- Cellini, S. R., & Goldin, C. (2014). Does federal student aid raise tuition? New evidence on for-profit colleges. *American Economic Journal: Economic Policy*, 6(4), 174-206.
- Cellini, S. R., & Koedel, C. (2017). The case for limiting federal student aid to for-profit colleges. *Journal of Policy Analysis and Management*, 36(4), 934-942.
- Cellini, S. R., & Turner, N. (2019). Gainfully employed? Assessing the employment and earnings of for-profit college students using administrative data. *Journal of Human Resources*, 54(2), 342-370.
- Chade, H., Lewis, G., & Smith, L. (2014). Student portfolios and the college admissions problem. *Review of Economic Studies*, 81(3), 971-1002.
- Chen, M. (2019). The value of US college education in global labor markets: Experimental evidence from China. Princeton University, Industrial Relations Section Working Papers, No. 627.
- Chen, M. Keith and Judith A. Chevalier. 2012. Are women overinvesting in education? Evidence from the medical profession. *Journal of Human Capital*, 6(2), 124-149.
- Chen, X., B. Elliott, S. Kinney, D. Cooney, J. Pretlow, M. Bryan, J. Wu, N.A. Ramirez, and T. Campbell (2019). Persistence, retention, and attainment of 2011-12 first-time beginning postsecondary students as of Spring 2017. First Look. NCES 2019-401. *National Center for Education Statistics*.
- Chetty, R., Friedman, J. N., Saez, E., Turner, N., & Yagan, D. (2017). Mobility report cards: The role of colleges in intergenerational mobility. NBER working paper, No. w23618.
- Chetty, R., Friedman, J. N., Saez, E., Turner, N., & Yagan, D. (2020). Income segregation and intergenerational mobility across colleges in the United States. *The Quarterly Journal of Economics*, 135(3), 1567-1633.
- Chetty, R., & Hendren, N. (2018). The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects. *The Quarterly Journal of Economics*, 133(3), 1107-1162.
- Clotfelter, C. T., Hemelt, S. W., & Ladd, H. F. (2018). Multifaceted aid for low-income students and college outcomes: Evidence from North Carolina. *Economic Inquiry* 56(1): 278-303.

- Cox, J. C., Kreisman, D. & Dynarski, S. (2020). Designed to fail: Effects of the default option and information complexity on student loan repayment. *Journal of Public Economics*, 192(1).
- Dadgar, M. & Trimble, M. J. (2015). Labor market returns to sub-baccalaureate credentials: How much does a community college degree or certificate pay? *Educational Evaluation and Policy Analysis*, 37(4), 399-418.
- Dale, S. B., & Krueger, A. B. (2002). Estimating the payoff to attending a more selective college: An application of selection on observables and unobservables. *The Quarterly Journal of Economics*, 117(4), 1491-1527.
- Dale, S. B., & Krueger, A. B. (2014). Estimating the effects of college characteristics over the career using administrative earnings data. *Journal of Human Resources*, 49(2), 323-358.
- Darolia, R., Koedel, C., Martorell, P., Wilson, K., & Perez-Arce, F. (2015). Do employers prefer workers who attend for-profit colleges? Evidence from a field experiment. *Journal of Policy Analysis and Management*, 34(4), 881-903.
- Daugherty, L., Martorell, P., & McFarlin, I. (2014). Percent plans, automatic admissions, and college outcomes. *IZA Journal of Labor Economics*, 3(1), 1-29.
- de Brey, C., Snyder, T. D., Zhang, A., & Dillow, S. A. (2021). Digest of Education Statistics 2019. NCES 2021-009. *National Center for Education Statistics*.
- Deming, D. J. & Noray, K. (2020). Earnings dynamics, changing job skills, and STEM careers. *Quarterly Journal of Economics*, 135(4), 1965-2005.
- Deming, D., & Dynarski, S. (2009). Into college, out of poverty? Policies to increase the postsecondary attainment of the poor. NBER working paper, No. 15387.
- Deming, D. J., Goldin, C., & Katz, L. F. (2012). The for-profit postsecondary school sector: Nimble critters or agile predators? *Journal of Economic Perspectives*, 26(1), 139-64.
- Deming, D. J., Yuchtman, N., Abulafi, A., Goldin, C., & Katz, L. F. (2016). The value of postsecondary credentials in the labor market: An experimental study. *American Economic Review*, 106(3), 778-806.
- Denice, P. (2015). Does it pay to attend a for-profit college? Vertical and horizontal stratification in higher education. *Social Science Research*, 52, 161-178.
- Denning, J.T., Eide, E., Mumford, K., Patterson, R.W., and Warnick, M. (Forthcoming). Why have college completion rates increased? An analysis of rising grades. *American Economic Journal: Applied Economics*.
- Deterding, N. M., & Pedulla, D. S. (2016). Educational authority in the “open door” marketplace: Labor market consequences of for-profit, nonprofit, and fictional educational credentials. *Sociology of Education*, 89(3), 155-170.
- Dillon, E. W., & Smith, J. A. (2017). Determinants of the match between student ability and college quality. *Journal of Labor Economics*, 35(1), 45-66.
- Dillon, E. W., & Smith, J. A. (2020). The consequences of academic match between students and colleges. *Journal of Human Resources*, 55(3), 767-808.
- DiNardo, J., Fortin, N. M., & Lemieux, T. (1996). Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. *Econometrica*, 1001-1044.

- Dynarski, S. M., Hemelt, S. W., & Hyman, J. M. (2015). The missing manual: Using National Student Clearinghouse data to track postsecondary outcomes. *Educational Evaluation and Policy Analysis* 37.1_suppl, 53S-79S.
- Dynarski, S. M. & Scott-Clayton J. (2013). Financial aid policy: Lessons from research in C. Rouse, L. Barrow, T. Brock (Eds.), *Future of children*, 23, The Trustees of Princeton University, Princeton, NJ.
- Dynarski, S. M. & Scott-Clayton, J. E. (2006). The cost of complexity in federal student aid: Lessons from optimal tax theory and behavioral economics. *National Tax Journal*, 59(2), 319-356.
- Dynarski, S., Libassi, C. J., Michelmore, K. & Owen, S. (2021). Closing the gap: The effect of reducing complexity and uncertainty in college pricing on the choices of low-income students. *American Economic Review*, 111(6), 1721-56.
- Evans, W.N., Kearney, M.S., Perry, B, & Sullivan, J.X. (2020). Increasing community college completion rates among low - income students: Evidence from a randomized controlled trial evaluation of a case - management intervention. *Journal of Policy Analysis and Management*, 39(4): 930-965.
- Firpo, S. (2007). Efficient semiparametric estimation of quantile treatment effects. *Econometrica*, 75(1), 259-276.
- Flores-Lagunes, A., & Light, A. (2004). Identifying sheepskin effects in the returns to education. *University of Arizona, memo*.
- Francis, A. M., & Tannuri-Pianto, M. (2012). Using Brazil's racial continuum to examine the short-term effects of affirmative action in higher education. *Journal of Human Resources*, 47(3), 754-784.
- French, R. & Oreopoulos, P. (2017). Behavioral barriers transitioning to college. *Labour Economics*, 47, 48-63.
- Fricke, H., Grogger, J., & Steinmayr, A. (2018). Exposure to academic fields and college major choice. *Economics of Education Review*, 64, 199-213.
- Fryer Jr, R. G., & Loury, G. C. (2005). Affirmative action and its mythology. *Journal of Economic Perspectives*, 19(3), 147-162.
- Fu, C. (2014). Equilibrium tuition, applications, admissions, and enrollment in the college market. *Journal of Political Economy*, 122(2), 225-281.
- Ge, S., Isaac, E., & Miller, A. (Forthcoming). Elite schools and opting in: Effects of college selectivity on career and marriage outcomes. *Journal of Labor Economics*.
- Goldin, C. (2014). A grand gender convergence: Its last chapter. *American Economic Review*, 104(4), 1091-1119.
- Goldin, C. & Katz, L. F. (2010). *The race between education and technology*. Harvard University Press: Cambridge, MA.
- Goldin, C., & Katz, L. F. (1999). The shaping of higher education: The formative years in the United States, 1890 to 1940. *Journal of Economic Perspectives*, 13(1), 37-62.
- González-Velosa, C., Rucci, G., Sarzosa, M., & Urzúa, S. (2015). Returns to higher education in Chile and Colombia. IDB working paper series, No. IDB-WP-587.

- Goodman, J., Gurantz, O., & Smith, J. (2020). Take two! SAT retaking and college enrollment gaps. *American Economic Journal: Economic Policy*, 12(2), 115-58.
- Goodman, J., Hurwitz, M., & Smith, J. (2017). Access to 4-year public colleges and degree completion. *Journal of Labor Economics*, 35(3), 829-867.
- Goodman, J., Hurwitz, M., Mulhern, C., & Smith, J. (2019). O brother, where start thou? Sibling spillovers in college enrollment. NBER working paper, No. w26502.
- Goodman, S. (2016). Learning from the test: Raising selective college enrollment by providing information. *Review of Economics and Statistics*, 98(4), 671-684.
- Gottlieb, J. D., Polyakova, M., Rinz, K., Shiplett, H., & Udalova, V. (2020). Who values human capitalists' human capital?: Healthcare spending and physician earnings. US Census Bureau, Center for Economic Studies.
- Grogger, J. & Eide, E. (1995). Changes in college skills and the rise in the college wage premium. *Journal of Human Resources*, 30(2), 280-310.
- Grove, W. A. & Hussey, A. (2011). Returns to field of study versus school quality: MBA selection on observed and unobserved heterogeneity. *Economic Inquiry*, 49(3), 730-749.
- Grove, W. A. & Hussey, A. (2014). Returns to MBA quality: Pecuniary and non-pecuniary returns to peers, faculty, and institution quality. *Labour Economics*, 26, 43-54.
- Gurantz, O. (2020). What does free community college buy? Early impacts from the Oregon Promise. *Journal of Policy Analysis and Management*, 39(1), 11-35.
- Gurantz, O., Howell, J., Hurwitz, M., Larson, C., Pender, M., & White, B. (2021). A national-level informational experiment to promote enrollment in selective colleges. *Journal of Policy Analysis and Management*, 40(2), 453-479.
- Gurantz, O., Pender, M., Mabel, Z., Larson, C., & Bettinger, E. (2020). Virtual advising for high-achieving high school students. *Economics of Education Review*, 75, 101974.
- Hamermesh, D. S. & Donald, S. G. (2008). The effect of college curriculum on earnings: An affinity identifier for non-ignorable non-response bias. *Journal of Econometrics*, 144(2), 479-491.
- Hanushek, E. A., Schwerdt, G., Woessmann, L., & Zhang, L. (2017). General education, vocational education, and labor-market outcomes over the lifecycle. *Journal of Human Resources*, 52(1), 48-87.
- Hastings, J. S., Neilson, C. A., & Zimmerman, S. D. (2013). Are some degrees worth more than others? Evidence from college admission cutoffs in Chile. NBER working paper, No. w19241.
- Heckman, J. J. (1998). Detecting discrimination. *Journal of Economic Perspectives*, 12(2), 101-116.
- Hendricks, L. & Leukhina, O. (2017). How risky is college investment? *Review of Economic Dynamics*, 26, 140-163.
- Hoekstra, M. (2009). The effect of attending the flagship state university on earnings: A discontinuity-based approach. *The Review of Economics and Statistics*, 91(4), 717-724.
- Holzer, H., & Neumark, D. (2000). Assessing affirmative action. *Journal of Economic literature*, 38(3), 483-568.
- Hoxby, C. M. (1997). How the changing market structure of U.S. higher education explains college tuition. NBER working paper, No. 6323.

- Hoxby, C. M. (2009). The changing selectivity of American colleges. *Journal of Economic perspectives*, 23(4), 95-118.
- Hoxby, C.M. (2015). Computing the value added of American postsecondary institutions. U.S. Treasury Report. Retrieved March 8, 2022, from <https://www.irs.gov/pub/irs-soi/15rpcompvalueaddpostsecondary.pdf>.
- Hoxby, C. M. & Avery, C. (2013). The missing ‘one-offs.’ The hidden supply of high-achieving, low-income students. *Brookings Papers on Economic Activity*, 2013(1), 1-65.
- Hoxby, C. M. & Turner, S. (2013). Expanding college opportunities for high-achieving, low income students. *Stanford Institute for Economic Policy Research Discussion Paper*, 12, 014.
- Hungerford, T. & Solon, G. (1987). Sheepskin effects in the returns to education. *The Review of Economics and Statistics*, 175-177.
- Hurwitz, M., Mbekeani, P. P., Nipson, M. M., & Page, L. C. (2017). Surprising ripple effects: How changing the SAT score-sending policy for low-income students impacts college access and success. *Educational Evaluation and Policy Analysis*, 39(1), 77-103.
- Hurwitz, M., Smith, J., Niu, S., & Howell, J. (2015). The Maine question: How is 4-year college enrollment affected by mandatory college entrance exams? *Educational Evaluation and Policy Analysis*, 37(1), 138-159.
- Hyman, J. (2017). ACT for all: The effect of mandatory college entrance exams on postsecondary attainment and choice. *Education Finance and Policy*, 12(3), 281-311.
- Hyman, J. (2020). Can light-touch college-going interventions make a difference? Evidence from a statewide experiment in Michigan. *Journal of Policy Analysis and Management*, 39(1), 159-190.
- Jacobson, L., LaLonde, R., and G. Sullivan, D. (2005). Estimating the returns to community college schooling for displaced workers. *Journal of Econometrics*, 125(1-2):271–304
- Jaeger, D. A., & Page, M. E. (1996). Degrees matter: New evidence on sheepskin effects in the returns to education. *The Review of Economics and Statistics*, 733-740.
- James, E., Alsalam, N. Conaty, J.C., & To, D. (1989). College quality and future earnings: Where should you send your child to college? *American Economic Review*, 79(2), 247-252.
- Jenkins, P. D., & Fink, J. (2016). Tracking transfer: New measures of institutional and state effectiveness in helping community college students attain bachelor’s degrees.
- Jepsen, C., Mueser, P., & Jeon, K.-S. (2016). The benefits of alternatives to conventional college: labor-market returns to proprietary schooling. IZA Discussion Paper No. 10007.
- Jepsen, C., Troske, K., & Coomes, P. (2014). The labor-market returns to community college degrees, diplomas, and certificates. *Journal of Labor Economics*, 32(1), 95-121.
- Jia, R., & Li, H. (2021). Just above the exam cutoff score: Elite college admission and wages in China. *Journal of Public Economics*, 196, 104371.
- Kane, T. J., & Rouse, C. E. (1995). Labor-Market Returns to Two-and Four-Year College. *American Economic Review*, 85(3), 600-614.
- Ketel, N., Leuven, E., Oosterbeek, H., & van der Klaauw, B. (2016). The returns to medical school: Evidence from admission lotteries. *American Economic Journal: Applied Economics*, 8(2), 225-54.

- Ketel, N., Leuven, E., Oosterbeek, H., & van Der Klaauw, B. (2019). Do Dutch dentists extract monopoly rents? *Journal of Health Economics*, 63, 145-158.
- Kim, T. (2021). Estimating pecuniary and non-pecuniary returns to college education for academically marginal students: Evidence from the college enrollment quota policy in south korea. *Economics of Education Review*, 83, 102142.
- Kinsler, J. & Pavan, R. (2015). The specificity of general human capital: Evidence from college major choice. *Journal of Labor Economics*, 33(4), 933-972.
- Kirkeboen, L. J., Leuven, E., & Mogstad, M. (2016). Field of study, earnings, and self-selection. *The Quarterly Journal of Economics*, 131(3), 1057-1111.
- Klasik, D. (2012). The college application gauntlet: A systematic analysis of the steps to four-year college enrollment. *Research in Higher Education*, 53(5), 506-549.
- Knott, J. H. & Payne, A. A. (2004). The impact of state governance structures on management and performance of public organizations: A study of higher education institutions. *Journal of Policy Analysis and Management*, 23(1), 13-30.
- Kozakowski (2020). Are four-year public colleges engines for mobility? Evidence from statewide admissions thresholds. Dissertation.
- Kreisman, D., Smith, J., & Arifin, B. (Forthcoming). Labor market signaling and the value of college: Evidence from resumes and the truth. *Journal of Human Resources*.
- Lang, K., & Weinstein, R. (2013). The wage effects of not-for-profit and for-profit certifications: Better data, somewhat different results. *Labour Economics*, 24, 230-243.
- Leigh, D. E. and A. M. Gill (2003). Do community colleges really divert students from earning bachelor's degrees? *Economics of Education Review*, 22(1), 23–30.
- Leighton, M. & Speer, J. D. (2020). Labor market returns to college major specificity. *European Economic Review*, 128, 103489.
- Li, H., Meng, L., Shi, X., & Wu, B. (2012). Does attending elite colleges pay in China? *Journal of Comparative Economics*, 40(1), 78-88.
- Light, A., & Strayer, W. (2000). Determinants of college completion: School quality or student ability?. *Journal of Human resources*, 299-332.
- Lindahl, L., & Regnér, H. (2005). College choice and subsequent earnings: Results using Swedish sibling data. *Scandinavian Journal of Economics*, 107(3), 437-457.
- Lindley, J., & Machin, S. (2016). The rising postgraduate wage premium. *Economica*, 83(330), 281-306.
- Liu, V. Y., Belfield, C. R., & Trimble, M. J. (2015). The medium-term labor market returns to community college awards: Evidence from North Carolina. *Economics of Education Review*, 44, 42-55.
- Liu, Y. T. & Belfield, C. (2014a). The labor market returns to for-profit education: Evidence for transfer students. CAPSEE Working Paper.
- Liu, Y. T. & Belfield, C. (2014b). Evaluating for-profit higher education: evidence from the education longitudinal study. CAPSEE Working Paper.
- Lochner, L. (2011). Nonproduction benefits of education: Crime, health, and good citizenship. In *Handbook of the Economics of Education* (Vol. 5, pp.183-282). Elsevier.
- Long, B. T. (2004). How have college decisions changed over time? An application of the conditional logistic choice model. *Journal of econometrics*, 121(1-2), 271-296.

- Long, B. T. and M. Kurlaender (2009). Do community colleges provide a viable pathway to a baccalaureate degree? *Educational Evaluation and Policy Analysis*, 31(1), 30–53.
- Long, M. C. (2008). College quality and early adult outcomes. *Economics of Education Review*, 27(5), 588-602.
- Long, M. C. (2010). Changes in the returns to education and college quality. *Economics of Education review*, 29(3): 338-347.
- Looney, A., & Yannelis, C. (2015). A crisis in student loans?: How changes in the characteristics of borrowers and in the institutions they attended contributed to rising loan defaults. *Brookings Papers on Economic Activity*, 2015(2), 1-89.
- Lovenheim, M. F. (2011). The effect of liquid housing wealth on college enrollment. *Journal of Labor Economics*, 29(4), 741-771.
- Lovenheim, M. F. & C. L. Reynolds. (2013). The effect of housing wealth on college choice: Evidence from the housing boom. *Journal of Human Resources*, 48(1), 1-35.
- Lovenheim, M. F., & Turner, S. E. (2017). *Economics of education*. Macmillan Higher Education.
- Lowry, R. C. (2001). Governmental structure, trustee selection, and public university prices and spending: Multiple means to similar ends. *American Journal of Political Science*, 45(4), 845-861.
- Ma, J., Pender, M., & Libassi, C.J. (2020). Trends in college pricing and student aid 2020. New York: College Board.
- MacLeod, W. B., Riehl, E., Saavedra, J. E., & Urquiola, M. (2017). The big sort: College reputation and labor market outcomes. *American Economic Journal: Applied Economics*, 9(3), 223-61.
- Marcotte, D. E., Bailey, T., Borkoski, C., & Kienzl, G. S. (2005). The returns of a community college education: Evidence from the National Education Longitudinal Survey. *Educational Evaluation and Policy Analysis*, 27(2), 157-175.
- Martorell, P. & McFarlin Jr., I. (2011). Help or hindrance? The effects of college remediation on academic and labor market outcomes. *Review of Economics and Statistics*, 93(2), 436-454.
- McLendon, M. K., Hearn, J. C., & Mokher, C. G. (2009). Partisans, professionals, and power: The role of political factors in state higher education funding. *The Journal of Higher Education*, 80(6), 686-713.
- Montgomery, M. & Powell, I. (2003). Does an advanced degree reduce the gender wage gap? Evidence from MBAs. *Industrial Relations: A Journal of Economy and Society*, 42(3), 396-418.
- Mountjoy, J. (Forthcoming). Community colleges and upward mobility. *American Economic Review*.
- Mountjoy, J., & Hickman, B. (2021). The returns to college (s): Estimating value-added and match effects in higher education. NBER working paper, No. 29276.
- Murphy, R., Scott-Clayton, J., & Wyness, G. (2019). The end of free college in England: Implications for enrolments, equity, and quality. *Economics of Education Review*, 71, 7-22.
- Oreopoulos, P. & Ford, R. (2019). Keeping college options open: A field experiment to help all high school seniors through the college application process. *Journal of Policy Analysis and Management*, 38(2), 426-454.

- Oreopoulos, P. & Petronijevic, U. (2018). Student coaching: How far can technology go? *Journal of Human Resources*, 53(2), 299-329.
- Oreopoulos, P., & Salvanes, K. G. (2011). Priceless: The nonpecuniary benefits of schooling. *Journal of Economic perspectives*, 25(1), 159-84.
- Ost, B., Pan, W., & Webber, D. (2018). The returns to college persistence for marginal students: Regression discontinuity evidence from university dismissal policies. *Journal of Labor Economics*, 36(3), 779-805.
- Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, 37(2), 187-204.
- Oyer, P. & Schaefer, S. (2019). The returns to elite degrees: The case of American lawyers. *Industrial and Labor Relations Review*, 72(2), 446-479.
- Page, L. C., & Scott-Clayton, J. (2016). Improving college access in the United States: Barriers and policy responses. *Economics of Education Review*, 51, 4-22.
- Page, L. C., Iriti, J. E., Lowry, D. J., & Anthony, A. M. (2019). The promise of place-based investment in postsecondary access and success: Investigating the impact of the Pittsburgh Promise. *Education Finance and Policy*, 14(4), 572-600.
- Page, L. C., Kehoe, S. S., Castleman, B. L., & Sahadewo, G. A. (2019). More than dollars for scholars: The impact of the Dell Scholars Program on college access, persistence, and degree attainment. *Journal of Human Resources*, 54(3), 683-725.
- Pallais, A. (2015). Small differences that matter: Mistakes in applying to college. *Journal of Labor Economics*, 33(2), 493-520.
- Pallais, A. & Turner, S. (2006). Opportunities for low-income students at top colleges and universities: Policy initiatives and the distribution of students. *National Tax Journal*, 59(2), 357-386.
- Pu, Y. (2020). College admission in three Chinese provinces: Province-specific versus pooling quotas. *China Economic Review*, 60, 101299.
- Reynolds, C. L. (2012). Where to attend? Estimating the effects of beginning college at a two-year institution. *Economics of Education review*, 31(4), 345-362.
- Rodríguez, J., Urzúa, S., & Reyes, L. (2016). Heterogeneous economic returns to post-secondary degrees: evidence from Chile. *Journal of Human Resources*, 51(2), 416-460.
- Rouse, C. E. (1995). Democratization or diversion? The effect of community colleges on educational attainment. *Journal of Business & Economic Statistics*, 13(2), 217-224.
- Rouse, C. E. (1998). Do two-year colleges increase overall educational attainment? Evidence from the states. *Journal of Policy Analysis and Management: The Journal of the Association for Public Policy Analysis and Management*, 17(4), 595-620.
- Saavedra, J. E. (2009). The learning and early labor market effects of college quality: A regression discontinuity analysis. *Investigaciones del ICFES*.
- Salto, D. J. (2014). Brazil: a for-profit giant. *International higher education*, (74), 21-22.
- Sander, R. & Bambauer, J. (2012). The secret of my success: How status, eliteness, and school performance shape legal careers. *Journal of Empirical Legal Studies*, 9(4), 893-930.
- Sasser, A. C. (2005). Gender differences in physician pay tradeoffs between career and family. *Journal of Human Resources*, 40(2), 477-504.

- Scott-Clayton, J. (2011). On money and motivation a quasi-experimental analysis of financial incentives for college achievement. *Journal of Human Resources*, 46(3), 614-646.
- Scott-Clayton, J. & Rodriguez, O. (2015). Development, discouragement, or diversion? New evidence on the effects of college remediation policy. *Education Finance and Policy*, 10(1), 4-45.
- Scott-Clayton, J., & Zafar, B. (2019). Financial aid, debt management, and socioeconomic outcomes: Post-college effects of merit-based aid. *Journal of Public Economics*, 170, 68-82
- Sekhri, S. (2020). Prestige matters: Wage premium and value addition in elite colleges. *American Economic Journal: Applied Economics*, 12(3), 207-25.
- Simkovic, M. & McIntyre, F. (2014). The economic value of a law degree. *The Journal of Legal Studies*, 43(2), 249-289.
- Sjoquist, D. L., & Winters, J. V. (2016). The effects of state merit aid programs on attendance at elite colleges. *Southern Economic Journal*, 83(2), 527-549.
- Smith, J. (2013). Ova and out: Using twins to estimate the educational returns to attending a selective college. *Economics of Education Review*, 36, 166-180.
- Smith, J., Pender, M., & Howell, J. (2013). The full extent of student-college academic undermatch. *Economics of Education Review*, 32, 247-261.
- Smith, J., Goodman, J., & Hurwitz, M. (2020). The economic impact of access to public four-year colleges. NBER working paper, No. 27177.
- Smith, J., & Stange, K. (2016). A new measure of college quality to study the effects of college sector and peers on degree attainment. *Education Finance and Policy*, 11(4), 369-403.
- Staiger, D., & Stock, J. H. (1997). Instrumental Variables Regression with Weak Instruments. *Econometrica*, 65(3), 557-586.
- Stange, K. (2012). Ability sorting and the importance of college quality to student achievement: Evidence from community colleges. *Education Finance and Policy*, 7(1), 74-105.
- Stange, K. (2012). An empirical investigation of the option value of college enrollment." *American Economic Journal: Applied Economics*, 4(1): 49-84.
- Stange, K. (2015). Differential pricing in undergraduate education: Effects on degree production by field. *Journal of Policy Analysis and Management*, 34(1), 107-135.
- Stevens, A. H., Kurlaender, M., & Grosz, M. (2019). Career technical education and labor market outcomes evidence from California community colleges. *Journal of Human Resources*, 54(4), 986-1036.
- Stinebrickner, R., & Stinebrickner, T. R. (2014). A major in science? Initial beliefs and final outcomes for college major and dropout. *Review of Economic Studies*, 81(1), 426-472.
- Sullivan, Z., Castleman, B., Lohner, G. & Bettinger, E. (2021). College advising at a national scale: Experimental evidence from the CollegePoint initiative. *EdWorkingPaper No. 19-123*.
- Svensson, L. G. (1999). Strategy-proof allocation of indivisible goods. *Social Choice and Welfare*, 16(4), 557-567.
- Tandberg, D. A. (2010). Politics, interest groups and state funding of public higher education. *Research in Higher Education*, 51, 416-450.

- Webber, D. A., & Ehrenberg, R. G. (2010). Do expenditures other than instructional expenditures affect graduation and persistence rates in American higher education? *Economics of Education Review*, 29(6), 947-958.
- Weiss, M. J., Ratledge, A., Sommo, C., & Gupta, H. (2019). Supporting community college students from start to degree completion: Long-term evidence from a randomized trial of CUNY's ASAP. *American Economic Journal: Applied Economics*, 11(3), 253-97.
- Winston, G. C. (1999). Subsidies, hierarchy and peers: The awkward economics of higher education. *Journal of Economic Perspectives*, 13(1): 13–36.
- Zimmerman, S. D. (2014). The returns to college admission for academically marginal students. *Journal of Labor Economics*, 32(4), 711–754.
- Zimmerman, S. D. (2019). Elite colleges and upward mobility to top jobs and top incomes. *American Economic Review*, 109(1), 1-47.