ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 19-70

Superficially Coupled Systems: The Organizational Production of Inequality in Higher Education

Christina Ciocca Eller
Columbia University

The rise of accountability standards has pressed higher education organizations to oversee the production and publication of data on student outcomes more closely than in the past. However, the most common measure of student outcomes, average bachelor's degree completion rates, potentially provides little information about the direct impacts of colleges and universities on student success. Extending scholarship in the new institutionalist tradition, I hypothesize that higher education organizations today exist as, "superficially coupled systems," where colleges closely oversee their technical outputs but where those technical outputs provide limited insight into the direct role of colleges and universities in producing them. I test this hypothesis using administrative data from the largest, public, urban university system in the United States together with fixed effects regression and entropy balancing techniques, allowing me to isolate organizational effects. My results provide evidence for superficial coupling, suggesting that inequality in college effectiveness exists both between colleges and within colleges, given students' racial background and family income. They also indicate that institutionalized norms surrounding accountability have backfired, enabling higher education organizations, and other bureaucratic organizations like them, to maintain legitimacy without identifying and addressing inequality.

VERSION: May 2019

Suggested citation: C. Ciocca Eller (2019). Superficially Coupled Systems: The Organizational Production of Inequality in Higher Education (EdWorkingPaper No.19-70). Retrieved from Annenberg Institute at Brown University: http://edworkingpapers.com/ai19-70

Superficially Coupled Systems: The Organizational Production of Inequality in Higher Education

Christina Ciocca Eller

Columbia University

May 25, 2019

Abstract

The rise of accountability standards has pressed higher education organizations to oversee the production and publication of data on student outcomes more closely than in the past. However, the most common measure of student outcomes, average bachelor's degree completion rates, potentially provides little information about the direct impacts of colleges and universities on student success. Extending scholarship in the new institutionalist tradition, I hypothesize that higher education organizations today exist as, "superficially coupled systems," where colleges closely oversee their technical outputs but where those technical outputs provide limited insight into the direct role of colleges and universities in producing them. I test this hypothesis using administrative data from a large, urban, public university system in the United States together with fixed effects regression and entropy balancing techniques, allowing me to isolate organizational effects. My results provide evidence for superficial coupling, suggesting that inequality in college effectiveness exists both between colleges and within colleges, given students' racial background and family income. I also find that certain college characteristics, such as college size, help to explain some of this inequality. Taken together, my findings suggest that institutionalized norms surrounding accountability have backfired, enabling higher education organizations, and other bureaucratic organizations like them, to maintain legitimacy without identifying and addressing inequality.

The research in this paper has been supported in part by a Spencer Foundation/National Academy of Education Dissertation Fellowship. Previous versions of this paper have been presented at Cornell University, Georgetown University, Harvard University, the University of California, Berkeley, the University of Chicago, the University of Notre Dame, and Yale University, as well as the annual meeting of the Association for the Study of Higher Education. I thank Thomas A. DiPrete, Amy Binder, Eric Grodsky, Shamus Khan, Alondra Nelson, Aaron Pallas and the participants at all prior presentations for their helpful comments and suggestions.

Keywords: accountability, organizational effects, higher education, college completion, race & social class

Corresponding Author: Christina Ciocca Eller, Columbia University Department of Sociology, 501 Knox Hall, 606 West 122nd Street, MC 9649, New York, NY 10027. Phone: (203) 520-9934. Fax: (212) 854-2963. Email: cmc2304@columbia.edu.

1 Introduction

Higher education is, and has been, the central cultural institution of the modern system. Over many centuries, it links an ever-expanding set of specific activities, roles, and organizations to a universal and unified cultural core. And it defines categories of certified persons as carrying these linkages and as possessing both the relevant cultural core and the specific authority and capacity to carry out the roles.
~ Meyer et al. 2007

Higher education serves as a stratifying agent in American society (Roksa et al. 2007). Though no longer exclusively the terrain of the country's elite, baccalaureate-granting colleges and universities continue to act as powerful gatekeepers. One of their primary functions is to sort entrants into the categories of "bachelor's degree (BA) recipients" and "dropouts" (Meyer 1977; Stevens et al. 2008). The stakes of this sorting process are undeniable: on average, BA recipients earn higher wages, accrue better health, and enjoy greater family stability than non-recipients (Hout 2012). As Meyer and colleagues (2007) suggest, BA recipients also reap more symbolic rewards, achieving enhanced social status, heightened authority, and recognition as the legitimate (or "certified") carriers of a society's dominant culture (**?**; Bourdieu and Passeron 1979). This broad array of benefits, together with the widespread, American belief in the link between individual merit, higher education, and upward social mobility (Grubb and Lazerson 2004; Hochschild and Scovronick 2004; Labaree 1997), have heralded an era where both policy makers and members of the general public have called for, "college for all" (Carnevale 2008; Obama 2009; Rosenbaum et al. 2017). Such calls persist despite abundant evidence that colleges and universities in the United States post relatively low BA completion rates when compared with those of peer countries: in the U.S., the average completion rate for full-time students within six years of college entry is 59 percent, whereas it is 76 percent in Germany, 79 percent in Denmark, and 85 percent in the United Kingdom (European Commission 2015, Table 4.3; Snyder et al. 2018, Table 326.10). Among students who do not complete a degree, traditionally underrepresented minority and lower-income students are vastly overrepresented (Bowen et al. 2009), injecting further inequality into an already stratified system.

Appreciating this gap between the meritocratic promise of BA completion and the reality of high and unequally distributed dropout rates, scholars have examined the form, function, and outcomes of colleges and universities in great depth (e.g. Arum and Roksa 2011; Bowen et al.

2009; Charles et al. 2009). Yet these efforts also have proved unequally distributed. Rather than examining the agentic role of colleges and universities in shaping observed student outcomes, most scholars have paid greater attention to student-level trajectories into and through college (Stevens et al. 2008).[1] This omission is consequential for two main reasons. First, colleges and universities are the distinct organizational units empowered to confer BAs, which, as discussed above, are one of the most important gateways to enhanced social and economic status. This function of higher education organizations is especially significant because BAs serve as a, "great equalizer," in American society: among those with terminal BAs, achieving the credential severs the otherwise strong and enduring link between social class origin and social class outcome (Hout 1988; Torche 2011).[2] Second, the rise in educational accountability practices in the United States has increased the visibility of colleges and universities as social actors with clear responsibilities towards the students they serve (Kelchen 2018). One such responsibility is to ensure adequate "returns on investment" in the form of degree completion; yet it is unclear to what extent publicized graduation rates and other well-recognized measures, such as college rankings, are a function of individual attributes and experiences prior to college entry or of the organizational practices of specific colleges (Arum and Roksa 2016; Raudenbush and Bryk 1986; Meyer 1997). Considering the centrality of colleges and universities to the broader American project of upward social mobility, deeper theorization and measurement of organizational impacts are required.

In this paper, I provide an empirical account of whether and to what extent public, baccalaureate-granting colleges independently affect students' BA completion outcomes, once pre-college student attributes and experiences are taken into account (hereafter, "college effects"). I particularly investigate whether college effects differ for minority or low-income students as compared with white or higher-income students, and whether the empirical isolation of

[1] More recently, scholars have started to bring together stratification-based, institutional, and cultural approaches to treat colleges and universities more holistically in their study of stratified student outcomes (Armstrong and Hamilton 2013; Binder and Wood 2013; Stevens 2009, 2015; Stuber 2012). However, this small group of scholars typically has drawn on case studies rather than large-scale data containing all students attending multiple colleges or universities to explore the contribution of higher education organizations to stratified student outcomes.

[2] There are two caveats to this statement. First, Zhou (2019) recently has produced evidence suggesting that a college degree does not heighten intergenerational income mobility among college graduates. However, the author bases this conclusion on a data re-weighting strategy that randomly allocates college completion to enrollees, constituting a rather implausible assumption that brings the results into question. Second, Torche (2011) has demonstrated the presence of a "U-shape" relationship between higher education and social origins. While the link between social class origins and outcomes is weak among those with a terminal BA, it reappears as much strong among those who go on to graduate education. The consequences of this pattern for stratification and inequality is an important and emerging area of research (Posselt and Grodsky 2017).

organizational effects leads to new information concerning college effectiveness. I do so using high-quality administrative data from the four-year and comprehensive colleges that comprise a large, urban, public university system in the United States: a context particularly well suited to the study of college effects for both practical and theoretical reasons. From a practical perspective, administrative data provide information on the entire population of individuals within a given organization, which is required for robust estimation of the aggregate effect of that organization on individual outcomes. From a theoretical perspective, public colleges and universities akin to those in my sample serve over 50 percent of the entire college-going population (Indiana University Center for Postsecondary Research n.d.). They also are an important context for traditionally underrepresented minority and low-income students, the groups typically of greatest interest in studies of educational inequality and mobility.

I situate this empirical analysis in a broader evaluation of how colleges and universities, as organizational actors, contribute to the process of stratification by race and income in contemporary American society. Existing research has demonstrated that admissions practices are one important lever through which colleges and universities shape the educational (and perhaps occupational) careers of potential matriculants (Grodsky 2007; Karabel 2006; Karen 1990; Stevens 2009), and that these practices often implicate specific organizational values, priorities, and constraints. I build on these findings to argue that colleges and universities additionally stratify students through the production of outcomes, and that these processes also are organizationally determined, though less visibly so. This argument provides a more contemporary response to the central question posed by Meyer and Rowan (1978:79) in their early, new institutionalist examination of educational organizations: why are large-scale educational bureaucracies able to leave organizational impacts on student outcomes "uncontrolled and uninspected"? Meyer and Rowan answer this question by postulating that the central purpose of schools is not to produce student outcomes, but rather to bestow ritual classifications upon curriculum, students, and teachers. By perpetuating such classifications, Meyer and Rowan argue, educational organizations gain public support, resources, and legitimacy. They also develop formally as "loosely coupled" systems (March and Olsen 1976; Meyer and Rowan 1977; Weick 1976), where those at the top of the organizational bureaucracy remain distant from technical work activities, and activities are further disconnected from outputs. In short, because classifications rather than outputs (which

Meyer and Rowan deem instruction) provide educational organizations with legitimacy, little oversight of outputs is required for organizational survival.

Yet the educational landscape has changed substantially since Meyer and Rowan (1977) devised their well-known argument. Specifically, the rise in educational accountability practices not only has increased the agency attributed to schools, but also has heightened public and political scrutiny of their impacts on student outcomes. In the higher education sector, for example, colleges and universities must report aggregated student outcome data to the U.S. government in order to receive federal funding support, while many public higher education organizations also must report outcomes to state legislatures to comply with "performance-based funding" schemes (Gándara and Rutherford 2018; Hillman et al. 2018). These formal changes have emerged alongside the rise of meritocracy as the primary, cultural logic structuring America's public school system, placing intensive demands on schools to ensure adequate intellectual and social growth for every, individual student regardless of family background (Labaree 1997). Given these circumstances, public educational organizations, in particular, must adjudicate today between a greater array of institutionalized expectations about their purpose and conduct in order to maintain their legitimacy.They also must operate in a more resource-constrained and competitive environment given the diminishing state and federal funds available for education, the increased demand for high-quality education, and the rise in popular though simplistic measures of organizational quality in addition to those imposed by federal and state governments, such as college rankings (Espeland and Sauder 2007, 2016). As a result, education organizations have begun to produce a substantial array of data to comply with legal and normative requirements surrounding accountability and quality.

My findings, however, suggest that this uptick in data production and publication in fact conveys very little accurate or comprehensive information concerning college effectiveness, whether in an absolute sense or in relation to foundational educational goals such as equity. Specifically, I find that substantial inconsistencies exist between the college effects I estimate and those measures produced using publicly available data. Once I measure organizational effectiveness using statistical strategies that account for student inputs and the college selection process, colleges that appear quite successful by virtue of their average graduation rates (the measure most commonly available to the public) actually appear to possess minimal, or even negative, effects on average

BA completion outcomes. The calculation of college effects by particular student groups further complicates the picture of effectiveness, as a college might prove very effective for particular groups (e.g. non-underrepresented or high-income students) and much less effective for others (e.g. traditionally underrepresented minority or low-income students). However, because public colleges produce the data required by state and/or federal law, as well as by popular media outlets such as *U.S. News & World Report*, they are able to uphold the formal imperative of accountability without explicitly demonstrating their impact on student success. This situation also implicates scholars and policy-makers, as it inhibits the production of knowledge that might usefully inform college- or system-level reforms.

In some respects, then, the governmental and public focus on accountability has backfired, focusing the time and attention of higher education personnel on broad, general measures rather than those capable of conveying college effectiveness. I hypothesize that this circumstance leads higher education organizations, and other organizations with similar bureaucratic structures and environmental constraints, to exist as, "superficially coupled systems." I use this term to describe organizations in which substantial bureaucratic oversight is directed towards the production of specific outputs, but where those outputs in fact convey very little information about the organization's actual effectiveness. In the following, I show that the largely cosmetic indicators of college quality produced as a result of institutionalized norms surrounding "accountability," "meritocracy," and "American individualism" both provide evidence for superficial coupling and conceal important information regarding the direct role of colleges and universities in producing unequal student outcomes.

2 The Contemporary Context of American Higher Education: Accountability and Individualism

Since the 1980s, both public and private actors increasingly have demanded that educational organizations be held accountable for students' scholastic outcomes. At the elementary and secondary levels, scholars have traced this growing emphasis on organizational accountability to the release of the federal report, *A Nation At Risk,* in 1983, which suggested that America's public schools were failing (Hallett 2010; Mehta 2013). In the higher education sector, scholars have cited

a wider array of forces leading to the increased focus on accountability, including rising costs in higher education, insufficient public funding to keep up with higher education expansion, concerns about student learning, the globalization of the higher education sector, and heightened reliance on (and trust in) quantitatively analyzed data (Huisman and Currie 2004; Kelchen 2018; Kurlaender et al. 2016).

Yet rather than constituting a singular or discrete reform policy, educational accountability instead has served as an all-encompassing, idealized logic of how schools can and should operate (Hallett 2010). The logic of accountability in turn structures action at multiple levels, shaping common understandings of organizational mission, identity, and priorities; spurring the creation of various organizational sub-units to comply with federally-mandated, state-required and/or publicly desired evidence of "college quality"; and guiding the standards used to measure faculty and student success (Arum and Roksa 2016; Kelchen 2018; Zemsky 2010). It also leads to real consequences for colleges and universities, since an inability to provide accountability data violates both state and/or federal law as well as social norms surrounding transparency and equitable access to information (Espeland and Sauder 2007, 2016; HEOA 2008). Due to its panoptic, deeply consequential character, accountability might be considered an "institutional logic" (Friedland and Alford 1991), or a, "socially constructed, historical [pattern] of material practices, assumptions, values, beliefs, and rules" that both structures and reconstitutes social reality (Thornton and Ocasio 1999: 804). By upholding the institutional logic of accountability, educational organizations maintain their public legitimacy.

But what, exactly, is required of higher education organizations to demonstrate compliance with accountability standards? One set of practices is inscribed in federal law as part of the Higher Education Act (HEA) of 1965 and its subsequent reauthorizations, which applies to all colleges and universities that receive federal funding. The most recent HEA reauthorization, signed into law in 2008, details a variety of reporting requirements, including admissions, enrollment, and graduation rates disaggregated by gender, race/ethnicity, and Pell Grant receipt; cost of attendance; and average financial aid received by students, among other information (HEOA 2008; NCES 2018). All of these data are reported annually and stored in a publicly accessible, though complex, database entitled the Integrated Postsecondary Education Data System (IPEDS). The National Center for Education Statistics (NCES) also uses this information to publish a vast array

of statistics annually, though this information is not attributed to individual colleges. In fact, the 2008 reauthorization of the HEA prohibits the federal collection of "unit records," or identifiable student-level data, ostensibly due to concerns over student privacy (Cowan and Kessler 2017; Ewell 2010). The results of this policy are twofold: first, no federal information exists (and therefore, none is publicized) tying college-level outcomes to individual cohorts of students; and second, state governments instead are responsible for overseeing the collection and evaluation of student-level record data. Results pertaining to college effectiveness therefore are produced piecemeal, state by state, if at all (Cunha and Miller 2014; Kurlaender et al. 2016).

Beyond the information required by federal law, many states have adopted performance-based funding (hereafter, "PF") schemes that tie funding for public colleges and universities to a variety of student- and organization-level outcomes. These schemes have emerged as a response to waning state funding for higher education in addition to the broader logic of accountability. While the precise outcomes and configurations of PF vary from state to state, all include metrics for degree completion and many incorporate credit accumulation, year-over-year retention, and research and service productivity (Li 2018; Shin 2010). Yet it is unclear to what extent state legislators seek to isolate higher education organizations' role in producing the outcomes used to measure college and university performance, especially because a relatively small number of states have begun to adjust performance metrics based on the student population served (Gándara and Rutherford 2018). Instead, more substantial evidence has mounted to suggest that PF strategies in their current form are not producing the intended effect of raising degree completion rates, despite the fact that 35 states have adopted PF schemes (Hillman et al. 2018). Instead, scholars have shown that some positive externalities (e.g. greater attention to data utilization and new efforts to strengthen student support policies and practices) are more common, though it remains unclear whether these shifts result in improved degree completion rates (Dougherty et al. 2016a,b). In short, similar to federal data reporting requirements, current formulations of state-level PF reporting do not appear to isolate the effects of higher education organizations or to systematically improve student outcomes.

A third external source of pressure on higher education organizations to report student outcome data comes from media outlets such as *U.S. News & World Report*, which compile and publish college rankings. Though these rankings are based on extremely subjective indicators of

college quality (Espeland and Sauder 2007; Zemsky 2010), they are easily digestible and widely available to the general public. The popularization of these rankings throughout the 1990s and 2000s has had sizable impacts on colleges and universities. As Espeland and Sauder (2007: 11) have demonstrated in their study of law school rankings, for example, administrators keep these rankings in mind when they, "define goals, assess progress, evaluate peers, admit students, recruit faculty, distribute scholarships, conduct placement surveys, adopt new programs, and create budgets." The authors go on to describe the ways that rankings serve as "self-fulfilling prophesies" in the Mertonian sense, or social constructions that shape behaviors that have the subsequent impact of reinforcing those social constructions, and as forces of commensuration, or transformations in cognition capable of shifting diverse qualities into more limited quantities that share a common metric (see also Espeland and Sauder 2016). In other words, rankings both directly impact organizational action and normalize quantification as central to their responsibilities as accountable actors.

Yet despite this tremendous increase in the generation and publication of data, relatively little data required by state governments or the federal government, or requested by news outlets such as *US News*, actually isolates the role of higher education organizations in producing student outcomes. Scholars have demonstrated that the measures calculated from both sources, as well as from voluntary, "opt-in" databases such as the National Survey of Student Engagement (NSSE), are largely cosmetic indicators of educational effectiveness (Kuh and Ikenberry 2009; Kurlaender et al. 2016). In the case of federally required data, the ban on collecting student record data means that most analysts cannot trace organizational impacts on student cohorts over time and correspondingly cannot apply rigorous statistical methods for isolating an organizational effect (Kelchen 2018; Meyer 1997; Raudenbush and Bryk 1986). In the case of data requested by news outlets, the information required, such as average graduation rates, college reputation, faculty resources, student selectivity, and alumni giving (Morse and Mason 2017), represents a rather simplistic conceptualization of college quality. Similar to the federal data, these media data typically are not based on specific student cohorts, leading to declarations of "college quality" based on limited statistical evidence. Even recent analyses ranking colleges based on their presumed contribution to upward social mobility do not directly isolate the college's role in

producing that mobility (Chetty et al. 2017).[3] In sum, though data apparently meant to demonstrate organizational effectiveness abound, nearly all publicly available information cannot and does not isolate the independent effect of colleges and universities on student outcomes, net of the characteristics and experiences students bring with them to college as well as the college selection process.

Contributing to this longstanding omission is the sharp shift in American culture towards viewing education as a private good, used by individuals to compete for desirable social positions, rather than as a public good used either to prepare students for active citizenship or to provide skills required to maintain national economic growth (Brint and Clotfelter 2016; Labaree 1997). The timing of this shift has paralleled that of the accountability movement, as the variety of social circumstances leading to calls for heightened accountability (eg. rising educational costs, scarcer federal funding, globalization) similarly have instigated a more individualistic conception of the purpose and value of education. Labaree (1997; 2012) has chronicled this transition in great depth, arguing that two important byproducts have resulted. The first is the reinforcement of classical liberal politics, based on personal liberty, free markets, and individual choice. The second is the rise of meritocracy as the guiding educational ideology, along with the accompanying understanding the opportunity, yet not outcomes, must be equal. The centrality of individual talent, effort, and compliance with institutional norms is notable in the shift towards individualism, as students are constructed as responsible for pursuing progressively higher levels of schooling to ensure upward social mobility – even if at the expense of other students. This "zero-sum competition" (Labaree 1997: 56) requires that educational organizations present themselves as able to, "organize the competition in a relatively fair and open manner." So, as long as the "competition" is perceived as fair, individual students (rather than the educational organizations they attend) can and should be held accountable for their own outcomes.

In placing the cultural accounts of accountability and "American individualism" side by side, a clear tension emerges. On one hand, colleges are asked to reveal their effectiveness in promoting student success. On the other, students are assumed responsible for their own success, especially if

[3]In their paper on the subject, Chetty et al. (2017) calculate the college contribution to student mobility without incorporating information on whether or not students have graduated a) at all or b) from the focal college. As a result, it remains unclear to what extent students' upward mobility may be attributed to their degree, their college, or their individual characteristics.

colleges are able to show that they offer meritocratic opportunities for students to succeed. This mixed message raises the question of whether higher education organizations experience any threat to their continued survival if they *do not* demonstrate their independent role in producing student outcomes. It also begs the question of whether, even if colleges and universities were required by legal or cultural imperatives to convey their effectiveness, analysts could draw upon appropriate, standardized statistical measures to provide rigorous estimates of school effectiveness. I examine these ideas below in the context of existing scholarly efforts to measure school effectiveness and to understand the unique contribution of educational organizations to social stratification in the U.S.

3 Educational Organizations and Inequality

What independent role do educational organizations play in shaping unequal student outcomes? Scholars have investigated this question at the elementary and secondary levels since the 1960s, when Congress commissioned sociologist, James Coleman, to study the extent of inequality in the country's public school system. Investigating school resources and student outcomes across 4,000 schools, Coleman and his team arrived at numerous findings related to school segregation, the influence of teachers and peers, and the geographic distribution of educational advantage and disadvantage (Coleman et al. 1966). Yet one of the most cited findings pertains to the explanation of variation in students' test performance. At the point of Coleman's study, differences between schools accounted for only a very small proportion of the overall test score variation observed, once socioeconomic characteristics were controlled. In fact, certain school resources, such as expenditures and facilities, appeared to matter relatively little (if at all) for students outcomes, a finding that still stands today (Alexander and Morgan 2016; Morgan and Jung 2016). In part due to this limitation, and in part because of the larger structural challenges and effects of school segregation, Coleman et al. ultimately concluded that black and other minority students had little opportunity to use schooling to "overcome deficiencies" produced by non-school factors (Coleman et al. 1966: 21). So, though certain school characteristics, such as excellent teachers, appeared particularly effective in raising (or lowering) the test performance of specific student groups, schools on the whole were seen as having little effect on overall inequality between dominant and non-dominant groups.

These findings have had far-reaching consequences for the subsequent study of schools (Downey and Condron 2016), shifting sociological attention away from investigations of the aggregate effects (or effectiveness) of schools and towards particular aspects of schools that may contribute to unequal student outcomes, such as tracking practices (Gamoran and Mare 1989; Lucas 2001; Oakes 1985), peer effects (Davies and Kandel 1981; Hallinan and Williams 1990; Riegle-Crumb et al. 2006), and teacher effects (Crosnoe et al. 2004; Downey and Pribesh 2004; Jennings and DiPrete 2010). The majority of studies that have focused on the quantification of school effects since 1966 have emerged from the perspective of educational economics and/or policy (e.g. Hanushek 1997; Heck 2000; Raudenbush and Willms 1995; Tekwe et al. 2004). These studies tend to define and test various statistical methods for quantifying school effects instead of applying these measures to study inequality or other substantive areas of interest. Recent sociological scholarship (Jennings et al. 2015; Legewie and DiPrete 2012) has focused more intensively on linking empirically rigorous calculations of school effects with inequality by race, social class, and gender, finding that, contrary to Coleman et al.'s (1966) findings, schools do, in fact, have direct impacts on performance gaps based on student background. However, these accounts do not include higher education organizations, nor do they consider the broader institutional contexts in which school effects are embedded.

Considering the relative dearth of scholarship connecting between-school effects with inequality at the elementary and secondary levels, it is not surprising that few sociologists have quantified these effects at the postsecondary level. Instead, as with the literature pertaining to elementary and secondary education, those studies that do focus on between-college effects emerge from an educational economics and policy perspective (Cunha and Miller 2014; Goodman et al. 2015; Hoekstra 2009; Kurlaender et al. 2016; Liu 2011). Among these studies, Cunha and Miller (2014) provide the most comprehensive assessment of how to measure college effects, adapting Dale and Krueger's (2002) analytical strategy for addressing non-random student assignment to colleges (described in greater detail below) and applying it to quantify college impacts. Using this approach, the authors find that the large discrepancies in average student graduation rates and early wages produced by "flagship" universities and regional public colleges in Texas decrease in size once pre-college student attributes and experiences are controlled. Yet some substantial between-college differences still are present, such as the nearly 20 or more percentage point gap in

the BA completion rates found between Texas A&M and non-flagship campuses such as the University of Houston downtown or Texas Southern University. In work focused on the two-year college context, Kurlaender et al. (2016) arrive at a similar conclusion, finding notable differences in the likelihood of associate's degree completion across California's public community colleges. While Cunha and Miller (2014) and Kurlaender et al. (2016) do not distinguish between the size of between- and within-college effects, nor do they delineate the effects they find by student race or income, their findings still suggest that more variation in student outcomes may be explained by between-college discrepancies than Coleman et al. (1966) found in the elementary and secondary contexts.

Sociological examinations of higher education provide greater insight into possible mechanisms driving these between-college differences, despite the lack of studies that quantify aggregate college-level effects. The mechanism that most sociological accounts examine is college quality. Building on research linking educational experiences with inequality, this literature investigates whether increases in college quality, usually proxied by college admissions rates (or "selectivity"), also elevate desirable student outcomes like BA completion once student background characteristics are held constant (Gerber and Cheung 2008). The evidence on this point is rather mixed, with some scholars finding that heightened college quality increases graduation rates (Bowen et al. 2009) and others find limited or no effects (Cunha and Miller 2009; Heil et al. 2014). Yet greater consensus exists around the advantages of high college quality for traditionally underrepresented minority students (Alon 2015; Bowen and Bok 2000; Bowen et al. 2009; Brand and Halaby 2006; Fischer and Massey 2007; Kurlaender and Grodsky 2013; Small and Winship 2007), with scholars determining that selective colleges elevate black and Hispanic students' graduation rates by 19 and 12 percent, respectively (Alon and Tienda 2005).

As recent work has noted, however, only a very small proportion of traditionally underrepresented students attends such colleges, limiting their overall effectiveness in mitigating educational gaps between traditionally advantaged and disadvantaged students (Ciocca Eller and DiPrete 2018). Also problematic is the fact that these studies rarely explore between-college distinctions other than college selectivity or within-college differences in student experiences to explain differences in student outcomes by race or class. Some recent research has shed light on the "black box" of college effects (Armstrong and Hamilton 2013; Arum and Roksa 2011; **?**;

Charles et al. 2009; Stevens et al. 2008; Stuber 2012), finding that organizational practices distinctly divide student experiences by social class and racial background. Yet the majority of these investigations offer small-scale examinations of a limited range of higher education organizations and a small number of students rather than estimating organizational impacts at the college population level. As such, the question of how differences between and within colleges independently shape distinct student outcomes requires further empirical investigation, especially when considering colleges of similar quality.

4 Institutional Explanations of the Structure and Function of Education Organizations

While the stratification tradition provides invaluable theoretical and empirical tools to study the relationship between education and inequality, it typically intervenes at the individual rather than the organizational level of analysis. It also rarely situates either individual or organizational action in the broader institutional context. For this reason, "new institutional" organization theory provides an important complementary perspective. According to DiMaggio and Powell (1991:8), who coined the term, "new institutionalism," this theoretical approach, "comprises a rejection of rational-actor models, an interest in institutions as an independent variable, a turn toward cognitive and cultural explanations, and an interest in properties of supraindividual units of analysis that cannot be reduced to aggregations or direct consequences of individuals' attributes or motives." In this way, new institutional theory envisions a fluid relationship between the environment and organizations, as environmental forces implicitly structure the lens through which individuals and organizations interpret reality and determine appropriate action. Both the cognitive and cultural "turns" in sociological theory (e.g. Berger and Luckmann 1967; Bourdieu 1977; Collins 1988; Garfinkel 1967; Goffman 1967) also are palpable in this approach, with neoinstitutional theorists emphasizing the taken-for-granted nature of practical reason, the embodiment of social structure, and the embeddedness of organizations and individuals in socially constructed scripts of acceptable behavior.

Largely because the structure and function of educational organizations contradicted then-well-established principles of organizational analysis (H.D. Meyer and Rowan 2006), schools

and colleges provided early neoinstitutionalists with important empirical fodder (DiMaggio and Powell 1983; Meyer 1977; Meyer and Rowan 1977, 1978; Meyer and Scott 1983; Scott and Meyer 1988). In the context of the present study, one of the most important theoretical perspectives developed using schools as the focal case concerns organizational structure and survival. When it was developed, this new institutionalist perspective presented an alternative to the then-prevalent assumption that organizations survived due to tight coupling between top and bottom levels of bureaucratic hierarchies, where the top levels maintained strict control over technical work activities and outputs produced by the bottom levels. New institutionalists instead built on ideas initially proposed by the organizational scholars, March and Olsen (1976) and Weick (1976), to argue that the survival of schools depended on the opposite, termed "loose coupling" (Meyer 1978; Meyer and Rowan 1977). In schools, these scholars observed, a clear disconnect existed between the administrative leadership (principals or presidents) and the technical work and outputs of the organization (teachers or faculty members producing instruction and in turn, student learning). While March and Olsen (1976) and Weick (1976) first used the term, "loosely coupled systems," Meyer and Rowan (1977) provided an explanation for its existence: schools depended on this disconnected organizational structure to reduce the tension between the explicit and implicit purpose of education, where the explicit (or technical/rational) purpose was to convey knowledge and produce learning but the implicit (or ceremonial/mythical) purpose was to manufacture social legitimacy. In part for this reason, Meyer and Rowan (1977) described schools as an ideal example of, "institutionalized organizations," or organizations primarily structured and constrained by the ceremonial imperatives of social legitimacy rather than that of technical efficiency.

The enmeshment of myths and rationality in Meyer and Rowan's (1977) work conveys a dynamic tension between two, discrete understandings of the social world, captured in divergent conceptualizations of what "institutions" actually are. From a "rational" standpoint, institutions might be viewed as a network of influential regulatory bodies such as state and professional agencies (Scott and Meyer 1983), while from a "cultural" (or "mythical") perspective, they might be conceived as, "repetitive social behavior underpinned by normative meanings and social understandings" (Greenwood et al. 2008). This dual character of institutions is in no way new to the sociological study of organizations, as the scholars who created the "old" institutionalism regularly described the tension between rationality and culture as a paradox. On the one hand,

organizations were, "formal structures subject to calculable manipulation," while on the other, they were, "inescapably embedded in an institutional matrix" (Selznick 1948: 25-26, quoted in Scott 2004). Early neo-institutionalists diverged from this perspective by viewing organizations not as embedded in institutions, but rather as the structural embodiment of cultural dynamics.

In more recent studies in the new institutional tradition, however, fewer scholars have attended to the complementary and perhaps even inseparable quality of rational or "regulatory," to use the word more often cited in the literature, and cultural institutional forces. As Greenwood et al. (2008) note, the divide between regulatory and cultural understandings of institutions has grown since the 1980s, despite efforts by Scott (2014) and others to demonstrate the concurrent operation of regulative, normative, and cultural-cognitive elements underlying institutions. In response to this division, a small group of scholars has begun to develop an approach that seeks to reunite rational and cultural understandings, known as the "inhabited institutions" (Hallett and Ventresca 2006) approach to organizational analysis (e.g. Armstrong and Hamilton 2013; Binder 2007; Hallett 2010; Stevens 2009). This approach re-emphasizes the role of people, social interactions, and meaning-making in the dual processes of responding to and creating institutionalized rules and rationalized myths.

As a particularly important example for the present study, Hallett (2010) describes the "recoupling" process between accountability practices and organizational action in a public elementary school, turning attention to the union of cultural and regulatory institutions in producing organizational action. Hallett demonstrates through this study that focusing attention on one or the other approach arguably would conceal vital information about organizational structure, priorities, and constraints, a principle equally relevant to the higher education context. For example, from a rational standpoint, maintaining legitimacy in higher education likely requires a college's demonstration that it properly uses public and private funds, and that it complies with governmental data reporting requirements. Meanwhile, from a cultural, more micro-level perspective, legitimacy more likely emerges from a college's capacity to demonstrate its central role in upholding important American values such as meritocracy, equity, and upward social mobility for individual students.

These two sets of "rationalized myths," to use Meyer and Rowan's (1977) phrase, represent the translation of the broader tension between accountability and American individualism to the

organizational level, as higher education organizations must act in ways that uphold both institutional logics in order to maintain social legitimacy. In some ways, the two institutional logics align with similar types of organizational action, as each requires that colleges publicly convey information about their effects on students. Yet the release of such information may pose a serious threat to the overall legitimacy of a college if it is contradictory in nature, conveying limited or even negative impacts on the likelihood of BA completion in the aggregate, and Horatio Alger-inspired success stories of students from modest backgrounds rising to greatness at the individual level. Considering that many colleges pursue these dualistic paths while maintaining legitimacy and continuing to operate, other dynamics must be at work.

Specifically, I argue that the present accountability regime – in which colleges are required to report unadjusted aggregates of students' BA completion outcomes rather than discrete measures of organizational impacts on BA completion – gives college and university leaders the opportunity to comply with accountability requirements without violating cultural expectations regarding the role of colleges and universities as vehicles for individual advancement. It does so by implicitly shifting the dialogue on college outcomes from one focused on *organizational* impacts to one focused on *individual* student outcomes, given the nature of the data reported. Cultural beliefs regarding the purpose of education in the United States, including individualism and meritocracy, support this logic, as success and failure in schooling ultimately is assumed to result from individual talent and effort rather than distinct organizational practices (Khan 2012; Labaree 1997, 2012).

At the organizational level, I further propose that the leaders of colleges and universities solve this clear tension between rational and cultural "rationalized myths" or "institutional logics" through a different kind of structural arrangement than that captured either by "loosely coupled systems" (Meyer and Rowan 1977) or "recoupled systems" (Hallett and Ventresca 2006), which I call, "superficially coupled systems." In such systems, bureaucratic leaders of higher education organizations exert apparently tight control over students' academic outcomes (e.g. technical outputs) through compliance with existing accountability reporting requirements, as shown by Espeland and Sauder (2007, 2016). However, the data they report potentially convey little information about their organization's role in producing those outcomes, again shifting the narrative from one of organizational impacts to that of individual ability and effort. College and

university leaders accordingly are able to sustain organizational legitimacy and survival; yet, largely due to the accountability regime, they also obscure the role of higher education organizations in producing unequal (meaning, non-meritocratic) student outcomes.

In the analyses below, I build on Espeland and Sauder's (2007, 2016) establishment of university leaders' clear responsiveness to the demands of accountability frameworks to address the question of whether that responsiveness, captured in publicly reported data, conveys adequate or accurate information about the direct role of colleges and universities in producing student outcomes. Together, these pieces of evidence should provide strong support for the logic of superficially coupled systems. I generate empirical evidence by comparing two sets of estimates of organizational effectiveness. The first set relies on measures included in public, federally reported, college data while the second incorporates substantially more student information, allowing me to statistically isolate the independent effect of colleges on student outcomes. These additional student-level data also facilitate the construction of college effects for particular student populations, including underrepresented minority students and low-income students. Drawing on the discussion above, I test four more specific hypotheses. First, statistical isolation of college effects should convey new and different information from estimates of college performance produced using publicly available data. Second, colleges should possess large, independent effects on student outcomes, above and beyond what might be expected based on students' pre-college attributes and experiences. In other words, once student characteristics are controlled, colleges still should influence students' likelihood of BA completion. Third, college effects should vary based on students' background, revealing substantial inequality in student outcomes both between and within colleges of similar "quality." Fourth, some college-level characteristics should help to explain the inequality demonstrated through the calculation of college effects, since colleges are implicated in the production of student outcomes.

5 Background and Data

I draw the primary data for this study from the administrative records collected on all entering students within the Metropolitan University (MetroU)[4] system between the fall semester of 2001

[4]I use pseudonyms throughout this work due to the terms of the data contract with the university system.

and the spring semester of 2008, which I received in de-identified form. MetroU is a large, urban, public university system in the United States. It is comprised of seven four-year (or, senior) colleges, four comprehensive (two- and four-year) colleges, and eight two-year (or, community) colleges and is predominantly a commuter system. Though the colleges maintain decentralized administrative structures, distinct histories and reputations, and disparate student bodies, they ultimately are affected by a similar array of institutionalized opportunities and constraints – especially because they are part of the same university system. In addition, among the four-year and comprehensive colleges, none is rated in the "most competitive" or "highly competitive" categories according to Barron's Selectivity Index, instead ranging from "competitive" to "noncompetitive" in their admissions standards for first-time freshmen.[5] Notably, however, a large proportion of the students attending any given MetroU four-year or comprehensive college represents two-to-four-year transfer students, who typically are admitted according to much more inclusive criteria than first-time freshmen. As a result, the MetroU colleges are more similar in their respective selectivity ratings than publicly available materials convey. At any given time, the MetroU system serves over 250,000 students predominantly from low-income backgrounds (more than 60 percent receive Pell Grants, requiring family income below $50,000) and traditionally underrepresented minority groups (more than 50 percent). About 40 percent of all students in the system are the first in their family to attend either a two-year or four-year college.

The data I received from MetroU provide a semester-by-semester account of students' academic trajectories while attending any college within the University system, including information on enrollment, credit accrual, major field of study, grade point average, and degree completion, among other attributes. These data also include matched student records with data from the National Student Clearinghouse (NSC), allowing me to track degree completion outcomes for students who have transferred out of the MetroU system to any other college within the U.S. This comprehensive information enables me to define college effects on BA completion in several different ways: first, as the effect of students' initial MetroU college on BA completion at any other college within or outside the MetroU system, second, as the effect of students' initial MetroU college on BA completion within the MetroU system, and third, as the effect of students'

[5] Of the eleven senior and comprehensive colleges, one possesses a Barron's Selectivity Index rating of 3, five possess a rating of 4, one possesses a rating of 5, and four possess a selectivity rating of 6, where lower ratings correspond with lesser selectivity. I consider the impacts of these ratings on the calculation of college effectiveness in Section 9.

initial MetroU college on BA completion within that same college. This latter measure is closest to that which is reported in federal data, yet it is also somewhat unrealistic given that thirty percent of students in the MetroU system transfer to another college at some point during their college journeys. I therefore produce estimates relying on all three measures but report within-college BA completion results to ensure comparability with existing reporting standards. In general, my results do not change substantially as a function of the outcome measure.

Considering that I have outcome data through to the spring of 2016, and that MetroU typically collects twenty semesters' worth of data for each student, I gain an eight- to ten-year period in which to observe student outcomes. This lengthy span of time is especially beneficial considering that most students in the U.S. require more than four years to complete a BA (Shapiro et al. 2014). I code the dependent variable in my analysis, BA completion, in three different ways to align with the three different definitions of completion described above. In all three of these scenarios, students are considered "non-completers" at the end of the twenty-semester window even if they still appear to be enrolled in college, though just 4 percent of the four-year beginning sample and 7 percent of the transfer student sample fits into this category (I discuss transfer students in greater depth below).[6] Based on the MetroU data I employ in the present analyses, the average BA completion rate for four-year beginning students who start college at an MetroU college and graduate at any U.S. college within 10 years of entry is 62 percent, while that figure is 56 percent if we evaluate graduation as a function of BA completion at any college in the MetroU system, and 49 percent if we estimate within-college completion rates.

Since BA completion outcomes are my primary focus, I first conduct analyses using the records of 81,985 students who begin their higher education experiences in the seven four-year colleges and four comprehensive colleges between 2001 and 2008. When including students from comprehensive colleges, I only incorporate those who have started in four-year programs. Across the MetroU system, the average college serves 7,464 four-year beginners. Table 1 presents the population of four-year beginning students in each college, separated by gender, racial background, family financial status, and transfer status (both inside and outside MetroU). I measure family financial status using an indicator in the data capturing whether or not students have received a federal Pell Grant. I use this indicator because the variable for student income is incomplete.

[6] As a sensitivity check, I omit this small group from my calculations, but the results are comparable.

Table 1: Proportional Distribution of Student Characteristics among Four-year Beginners within MetroU Colleges (*n=81,985)*

	Total Number of Students	Identifies as Male	Identifies as White	Identifies as Black	Identifies as Hispanic	Identifies as Asian	Pell Grant Recipient	Transfer within MetroU	Transfer Outside of MetroU
College 1	12,122	0.505	0.348	0.107	0.171	0.374	0.480	0.061	0.105
College 2	11,068	0.400	0.481	0.079	0.175	0.265	0.385	0.158	0.139
College 3	2,428	0.441	0.759	0.041	0.105	0.094	0.247	0.154	0.157
College 4	9,858	0.437	0.472	0.224	0.119	0.185	0.501	0.231	0.133
College 5	13,702	0.333	0.400	0.146	0.208	0.246	0.427	0.236	0.182
College 6	7,798	0.411	0.323	0.205	0.389	0.083	0.536	0.221	0.186
College 7	9,731	0.527	0.1502	0.259	0.353	0.238	0.537	0.258	0.151
College 8	6,536	0.314	0.077	0.297	0.564	0.062	0.687	0.203	0.176
College 9	6,195	0.351	0.057	0.567	0.225	0.151	0.604	0.322	0.178
College 10	1,691	0.737	0.138	0.358	0.353	0.151	0.555	0.278	0.155
College 11	856	0.331	0.007	0.899	0.081	0.013	0.568	0.255	0.196

Source: MetroU administrative data.

However, Pell Grant status, reported via federal financial aid forms, clearly demonstrates whether a student's family income is above or below $50,000, providing useful information pertaining to students' financial status.

In addition to the population of four-year beginning students, the MetroU system data also incorporates information for students who have started in a two-year program at an MetroU college and who have subsequently transferred to a four-year MetroU program or college. As with many public colleges, the transfer population is large at MetroU: at any given point, two-thirds of MetroU's total student population attending four-year colleges is comprised of transfer students and transfer students additionally obtain roughly 60 percent of all MetroU bachelor's degrees. However, empirical accounts of transfer students' experiences and outcomes are limited for a variety of reasons, including data constraints, analytical difficulties resulting from transfer students' multiple college selection processes, and theoretical challenges given substantial differences in transfer students' family background and educational preparation, as compared with four-year beginning students (see Ciocca Eller and DiPrete 2018, Online Appendix B for an elaboration of these issues). Though the latter two challenges remain in the present study, I

analyze college effects separately for the distinct pool of 44,470 students who transfer from two-year to four-year colleges within the MetroU system, while also comparing college effects estimates for transfer students with those of four-year beginning students in Online Appendix A. I do so because of the large proportion of two-to-four-year transfer students that comprises the four-year MetroU system[7] as well as the importance of the transfer pathway to traditionally underrepresented students who seek to earn a BA (Bailey et al. 2015; Bowen et al. 2009; Fink and Jenkins 2017; Jenkins, Davis and Fink, John 2016). That said, for analytical clarity, I focus on the results for four-year beginners in the main body of the text.

Beyond chronicling the academic trajectories of MetroU students once they arrive in college, the administrative data also include extensive pre-college information collected from students' applications and, when relevant, FAFSA forms. From the application data, I am able to gather information pertaining to students' family background, pre-college academic achievement, economic standing, and high school characteristics. I also can observe whether students have been offered financial aid prior to college entry, an important consideration for lower-income students, in particular. I standardize all linear pre-college measures, including high school grade point average and math and English state test scores, accumulated college credits through advanced placement or international baccalaureate classes, students' total financial aid award granted prior to college entry, and family income, so that they have a mean of 0 and a standard deviation of 1. In all analyses, I additionally incorporate one college-level variable, the proportion of students who transfer out of the college, as its inclusion facilitates unbiased comparison of college success for those students who remain within the college population. Finally, for analyses in Section 9, I define and incorporate a number of additional college-level variables, including college size, percent underrepresented minority, percent Pell, and a binary indicator for "mid-selective" versus "non-selective." All variables employed in my analyses are summarized in Appendix Table 1.

Unlike the data describing students' academic trajectories while enrolled in college, which are complete, the pre-college information includes some missing information. Most variables have a relatively low level of missingness, about 10 percent or less. However, some other variables, such

[7]Transfer between different four-year campuses also is a common path within the MetroU system. On the whole, 30 percent of MetroU students start at one MetroU college and transfer to another. So, while the population of four-to-four-year transfers on MetroU colleges' campuses is not inconsequential, it is nowhere near as large as the population of two-to-four-year transfer students. For this reason, I focus on the two-to-four-year transfer population in this paper rather than the four-to-four-year transfer population.

as students' high school test scores, have higher missingness of closer to 25 percent. I generally handle missing data using multiple imputation with chained equations (Allison 2002; Peugh and Enders 2004; Rubin 1976), a strategy based on the assumption that data are missing at random, conditional on all the other covariates in the imputation model. Based on the extensive student data available through the MetroU system, this assumption is fair for all pre-college variables save for one, student income. Based on the collection procedures MetroU used, student income data is completely missing for several incoming cohorts, including 2001, 2003, and 2005. I handle this non-random missing data in several ways: first, I calculate results without incorporating income data, relying instead on coarser measures of income status such as Pell grant information; second, I restrict my analysis to those years without completely missing income data; and third, I impute income data using a model pooled across all available years of data. Regardless of the strategy, my results remain comparable. Accordingly, I incorporate the imputed income data because it slightly improves the statistical power of my estimates.

6 Methods

Scholars primarily have pursued two analytical strategies to measure the independent effect of schools on student outcomes, random effects regression (Jennings et al. 2015; Kurlaender et al. 2016; Meyer 1997; Raudenbush and Bryk 1986; Raudenbush and Willms 1995) and fixed effects regression (Cunha and Miller 2014; Dale and Krueger 2002, 2014). The random effects approach defines the "school effect" as the standard deviation of schools' average effect on a given outcome, after accounting for student-level differences. As Raudenbush and Bryk (1986) put it in their seminal work on the approach, it is a "slopes as outcomes" operationalization. In contrast, the fixed effects approach measures the school effect as the coefficient on an indicator variable, which is entered into regression models alongside extensive controls for student background and the college selection process. Despite formal differences, the goal of both types of models is to statistically isolate the effect of schools by relying on a counterfactual, or "potential outcomes" framework (Angrist and Pischke 2009; Morgan and Winship 2007; Rubin 2005). In this framework, individuals are envisioned to possess two outcomes: that which is actually observed given a particular "treatment" (in this case, entering a specific school), and that which

hypothetically could occur if individuals were exposed to a different treatment (attending a different school). In this sense, both approaches seek to identify the "value added" by the school, or that portion of the overall effect that is caused by school characteristics and practices rather than individual-level traits or contextual circumstances.

In the present study, I employ a fixed effects strategy due to several advantages it has over the random effects approach. The most important of these is that fixed effects models more adequately control for unmeasured characteristics at the college level, since each college is treated as its own, unique context in which to examine the relationship between enrolled students and outcomes of interest. That said, fixed effects models also can be limiting in some instances; for example, they rest on the assumption that all possible confounders with time-variant values or effects have been incorporated into the estimation model. They also require that adequate variability in covariates and responses within each "fixed" group is present to produce unbiased results (Rabe-Hesketh and Skrondal 2012). Due to the very large number of students with variable characteristics and outcomes in my sample, the requirement regarding variability in covariates is easily met. Regarding the issue of variance over time, it is possible that college characteristics might change and therefore introduce bias into the estimation of college effects. Yet this source of bias at the college level is less likely than bias emerging from unmeasured characteristics at the student level.

I address the possibility of bias at the student level by incorporating a large array of student-level covariates as discussed in the "Background and Methods" section. I also employ a complementary statistical strategy that powerfully addresses the possibility of omitted variables, the "within-matched-applicant" (WMA) approach. I estimate WMA fixed effects models using logistic regression because of the binary outcome variable of BA completion. These models take the form:

$$logit[P(BA = 1)] = \beta_1 + \beta_2 X_i + \beta_3 G_i + \beta_4 C_i + \beta_5 T_i + \beta_6 O_s \quad (1)$$

where X_i is a vector of individual characteristics captured prior to college entry, G_i is an indicator for students' application/admissions group, C_i is an indicator for enrollment in one of the MetroU colleges as compared to any of the others, T_i is a fixed effect for year and month of college entry,

and O_s is the average transfer-out rate of each college in the MetroU system. Dale and Krueger (2002), who proposed the WMA method, argued that this analytical strategy would reduce the effects of selection bias because of G_i. This parameter indicates students' inclusion in a "matched applicant" group based on application preferences and admissions decisions, where students are grouped together if they have applied to the same set of colleges and received the same combination of acceptances and rejections from those colleges. According to Dale and Krueger, this term should mitigate bias so long as, a) admissions decisions reveal more about students than characteristics already observed in any given data set, and b) the combination of application and admissions decisions capture observable and unobservable similarities and differences across students that would not be revealed without these preferences and decisions. It follows that Equation 1 should produce an unbiased estimation of the college effect in the coefficient, β_4, which is the increase or decrease in the log odds of BA completion associated with attending a particular MetroU college after controlling for student-level characteristics, students' application preferences and admissions decisions, the year and month of enrollment, and the average transfer-out rate of the colleges in the MetroU system.

Though Equation 1 follows Dale and Krueger's (2002) formulation of WMA closely, I make two adjustments. First, Dale and Krueger's matching approach relies on the number of colleges to which a study has applied together with the average SAT score of those colleges. However, using the MetroU administrative data, I gain more granular information. This information allows me to group students together based on the total number of MetroU colleges to which they have submitted applications (as low as one and as high as six), the exact colleges to which they have applied, and their preference ordering among the colleges, along with the admissions decisions accompanying each application. Within the four-year beginner sample, I end up with 4,149 groups of matched applicants with two or more students per group; the mean size of each group is 48 students. I then enter each of these groups into the fixed effects model as indicators, omitting Group 1. While I do not have data for students' college applications outside of the MetroU system, the order of students' college preferences, together with their high school academic performance, provide a strong indication of their self-perceived and objectively recorded preparedness for college success.

Second, existing studies of college effects employing the WMA method (e.g. Cunha and

Miller 2014) select a college to omit from the equation, contrasting the effectiveness of that college with the performance of each of the other colleges in the system. In other words, the counterfactual in these models is students' attendance of a particular, omitted college as compared with the college that students actually attend. In my approach, however, I typically estimate eleven, separate models in which each college serves as the focal "treatment" and the counterfactual is not students' BA completion rate in any particular college, but rather students' rate if they were to attend an average "other" college within the MetroU system. I choose this approach because it allows me to calculate effects for every college in the system in relation to all of the other colleges instead of just one college, providing a more expansive and parsimonious understanding of "effectiveness." For analyses in which I do calculate pairwise comparisons of college effects, contrasting, for example, the effectiveness of College 1 versus Colleges 2 through 11, I still calculate ten separate regression models. In this scenario, College 1 serves as the treatment and Colleges 2 through 11 separately serve as the the control. I pursue this additional step because it enables an understanding of the extent of between-college differences in student BA completion outcomes across the entire MetroU system.

In sum, my approach generates estimates of between-college effectiveness using separate logistic regression models where the college effect is captured in the coefficient, β_4. For ease of interpretation, I subsequently employ predictive margins to estimate the BA completion rate for students who attend the focal college, as well as the counterfactual graduation rate that would emerge if these *same students* where to attend the average, non-focal college.[8] I then difference these two rates, calculating the percentage point increase or decrease in BA completion rates incurred among the students who actually attend each of the MetroU colleges. This difference again represents the between-college effect, though in a transformed (and more interpretable) form. As a final step, I employ chi-square tests to test for differences between the actual and counterfactual rates produced by the predictive margins.

Despite the usefulness of the WMA approach, it presents analytical challenges that must be addressed. Namely, it provides no additional information for students who apply to just one college (16 percent of four-year beginners in this sample), for whom application records are

[8] I also have produced these results by defining the counterfactual as the average outcome across all colleges in the MetroU system. The results are nearly identical because the treatment and control samples are well balanced due to the ENT and WMA methods.

missing (7 percent of four-year beginners), or for whom there is no match in the data for the exact combination of six ordered college preferences and six admissions decisions (49 percent) (Dale and Krueger 2002). Additionally, in the case of binary dependent variables (such as those required to study BA completion), entire matched-applicant groupings are omitted from the equation if they achieve the same outcome. It is also possible that the WMA approach is less revealing about MetroU students' unobserved characteristics than when employed in Dale and Krueger's (2002, 2014) studies of more selective colleges, especially due to the more compressed distribution of college selectivity among MetroU colleges. In this case, selection bias would remain in the WMA results despite my various adjustments.

I address the challenges posed by lack of additional information for various groups by creating separate, "matched" groupings for students who have applied to zero colleges, students for whom application records are missing, students with no match in the data, and students who would be dropped due to common outcomes. I do so because it seems likely that unobserved similarities are present among students based on the number of applications they submit, the uniqueness of their preferences, and among students with the same application and acceptance profile, even if they do experience the same college outcome. I ensure that this approach is robust in light of the high proportion of students with no exact WMA grouping through two sensitivity analyses, the first of which uses a slightly coarser matching method that results in less than one percent of students without a direct match[9] and the second of which omits all unmatched students. The results from both of these test are comparable to those produced when using the additional matched groups. I therefore proceed using all actual matches in addition to the manufactured matches given lack of application information, application to only one school, lack of exact match, and common outcomes in matched groups, since it optimizes granular application data for 30 percent of the sample.

I address the possibility that the WMA method does a poorer job of correcting for selection bias than in Dale and Krueger's (2002, 2014) studies by turning to an additional, and quite different, analytical approach, entropy balancing (ENT). ENT is a generalization of the conventional propensity score matching approach, in which student-level observations are weighted

[9]This approach matches students on three application preferences and admissions decisions rather than six, even if students have submitted more than three applications.

based on their observable likelihood of selecting into a given treatment condition (Rosenbaum and Rubin 1983). This reweighting procedure mitigates the adverse effects of selection bias on analytical results since students in treatment and control groups are "balanced" across observable characteristics. Entropy balancing is similar to propensity score matching in that it assigns weights to cases based on an individual's likelihood of selecting into treatment. However, it also makes use of an analyst's knowledge of various statistical moments of the covariate distribution, including mean, variance, and skewness, to calculate weights that will ensure exact matching across treatment and control groups according to these moments (Hainmueller 2012). This approach is advantageous because it is "doubly robust" (Hainmueller and Xu 2013; Hirano and Imbens 2001; Zhao and Percival 2017), meaning that statistically accurate results will emerge even if the analyst only properly specifies one of the two required equations, the treatment equation and the outcome equation.

Accordingly, I estimate eleven separate models using the ENT approach, one for each MetroU college. In the first step, I delineate a focal college as the treatment and all other colleges as the control, akin to my approach for the WMA estimates. I then balance treatment and control groups by matching students based on the mean, variance, and skewness of all pre-college covariates for the treatment group, as described formally by Hainmueller (2012).[10] Once the treatment and controls groups are balanced, I subsequently employ fixed effects logistic regression as in Equation 1, with two exceptions. First, I do not include the WMA groupings in this estimation due to the redundancy between the WMA groups and the ENT reweighting strategies, and second, I incorporate the newly defined balancing weight in the subsequent regressions. I then follow the procedure detailed above for the WMA approach, predicting linear margins for BA completion in the actual and counterfactual states, differencing the two, and conducting chi-square tests on the marginal values. Finally, I compare the WMA and ENT results to ensure stability of the college effects I calculate. While I employ both the WMA and ENT approaches to calculate all college effects described below, I typically present the ENT results in my tables because of the

[10]In practice, this approach to balancing consists of three main features: a loss function based on Kullback's (1959) entropy divergence, which measures the distance between the distribution of estimated control and treatment weights; a balancing constraint imposed by the researcher to determine the tolerance for difference between the mean, variance, and skewness of each covariate used to balance control and treatment observations; and two normalization constraints ensuring that the weights sum to the normalization constant of one and that the weights are nonnegative, as the distance metric is not defined for negative values (Hainmueller 2012: 31).

greater statistical robustness they provide.

The procedures described above reflect my approach to computing average, between-college effects. Yet I also study college effectiveness as a function of students' racial background and family income, both between and within colleges. To generate between-college estimates, I use the same WMA and ENT analytical models described above but restrict them to separate racial groups, including white, black, Hispanic, and Asian, and to separate income groups, including non-Pell and Pell. This approach produces "within-race/income, between-college" estimates of college effectiveness, showing whether a certain college is more or less effective for the students who comprise a particular racial or economic group. I also produce "within-college, between-race/income" results to make direct comparisons regarding the effectiveness of each MetroU college on one racial or income group versus another. To account for covariate imbalances between students of different racial and income groups for these within-college estimates, I again use the ENT technique. However, whereas each college served as the treatment in the between-college effects, students of particular racial and income groups act as the treatment in the within-college models. Specifically, each non-white student group serves as a unique treatment condition in the case of racial comparisons and Pell students act as the unique treatment condition in the income comparison. I also remove financial variables other than Pell receipt, such as family income and the receipt of state grants, from the Pell and non-Pell models in order to concentrate the effects of income into this single variable. For the within-college models, then, the college effect represents the difference between the BA completion rates of black, Hispanic, and Asian students as compared with white students and Pell students and compared with non-Pell students.

As one final step, I examine whether a variety of college-level variables, including college size, percent underrepresented minority, percent Pell, and college selectivity, predict BA completion and therefore explain the existing college effects.Here, fixed effects regression is not appropriate because college-level characteristics are uniform within each college grouping and therefore drop out of the estimation model due to multiple collinearity. I therefore draw on mixed effects logistic regression, inserting each college into the model as a random effect. This strategy allows me to isolate the impact of college-level variables on BA completion while holding both individual-level variables and colleges constant. Notably, I am somewhat limited in the use of administrative data to explain existing college effects due to the small number of colleges in the

MetroU system, which leads to relatively few degrees of freedom in regression analyses and in turn, restrictions on the number of explanatory variables it is possible to incorporate into analytical models. Though future research should test mechanisms using data from larger university systems possessing greater internal variation, I address the existing statistical limitation by incorporating insights from longitudinal interview data I collected with sixty students attending three, MetroU colleges. A comprehensive account of these data is available in Ciocca Eller (2019).

7 Providing a Foundation for the Estimation of College Effects

Despite the lack of publicly available student-level data required to calculate college effects, colleges do report some measures of student outcomes annually, whether to federal agencies, opt-in data collection projects, or media outlets. The most applicable of these measures is the within-college rate of graduation for first-time college entrants. To be clear, these rates count students who have transferred out of the specific college as non-completers, regardless of whether they have received a BA at a different college in subsequent years. Table 2 presents this within-college information for each college in the MetroU system, showing the average BA completion rate for all entry cohorts between 2001 and 2008. I include both eight-year and ten-year completion rates to demonstrate the usefulness of expanding the window of observation beyond what is publicly available. As a point of comparison, I additionally incorporate completion rates using the most lenient definition of BA completion, or completion at any college within or outside of the MetroU network. I do so because these numbers are more accurate representations of BA completion rates given the sizable proportion of students that transfer out of the MetroU system. I number each college according to its point estimate on the within-college BA completion metric, so that College 1 posts the highest graduation rate and College 11 the lowest.

As Table 2 conveys, substantial between- and within-college variation exists in the MetroU system. Between colleges, the difference in BA completion rates spans nearly 40 percentage points, as the ten-year, within-college BA completion rate in College 1 is 64.5 percent while it is 19.9 percent in College 11. Within-college variation is even more sizable; in most MetroU colleges, the proportion of degree recipients and non-recipients is about equal. These statistics suggest that, just as Coleman et al. (1966) found, within-school differences in student outcomes are larger than

Table 2: BA Completion Rates for Enrollees in MetroU's Senior and Comprehensive Colleges among Four-year Beginners

	Eight-year Within-College BA Rate (%)	Ten-year Within-College BA Rate (%)	Average Within-College Ten-year BA Completion Rate across All MetroU Colleges (%)	College Ranking Using Within-College Ten-year BA Rate	Eight-year Any College BA Rate (%)	Ten-year Any College BA Rate (%)	Average Any College Ten-year BA Completion Rate across All MetroU Colleges (%)	College Ranking Using Any College Ten-year BA Rate
College 1	63.7	64.5	48.7	1	74.0	76.7	61.9	1
College 2	56.0	57.1	48.7	2	65.4	68.5	61.9	2
College 3	51.0	52.3	48.7	3	62.5	65.5	61.9	3
College 4	50.2	51.1	48.7	3	60.0	63.5	61.9	5
College 5	46.9	48.0	48.7	5	60.6	64.9	61.9	3
College 6	44.0	45.2	48.7	6	54.3	57.5	61.9	6
College 7	43.7	44.9	48.7	6	55.5	60.0	61.9	7
College 8	38.5	39.9	48.7	8	46.4	50.1	61.9	8
College 9	27.5	28.9	48.7	9	37.9	42.7	61.9	9
College 10	23.4	24.7	48.7	10	32.2	36.7	61.9	10
College 11	18.9	19.9	48.7	11	28.7	31.2	61.9	11

Note: Rankings are constructed based on the presence of statistically significant differences in the mean BA completion values of each college using pairwise tests. Non-significant differences are signified by "ties." So, the average, eight-year within-college BA completion rate in College 2 differs significantly from that posted in College 1, but the difference between College 3 and College 4 is not statistically significant.
Source: MetroU administrative data.

between-school differences. According to calculations in which I compare the r-squared values from linear probability models without and with covariates (0.041 and 0.131, respectively; full models available upon request), between-college differences account for about 30 percent of the total explained variance while within-college differences explain the remaining 70 percent. Linear probability models are an appropriate approach to calculate these statistics because the overall completion rate is nearly 50 percent, coinciding with the most linear area of the logistic curve.

Beyond demonstrating the presence of between- and within-college differences in BA completion rates, Table 2 also points to two additional findings. First, not all BA completion rates are statistically different from one another, and second, the relative performance of colleges in the within-college and any-college conditions are actually quite similar. When evaluating the ten-year, within-college BA rates, Colleges 3 and 4 appear statistically equivalent, as do Colleges 6 and 7. When assessing the ten-year BA completion rates across all colleges, Colleges 3 and 5 are statistically equal, as are Colleges 6 and 7 once again. The main difference between the average completion rates when evaluating within-college and any-college rates is the greater proportion of completers in the any-college condition, an expected trend given the non-trivial proportion of transfer students. In addition, regardless of which definition of BA completion I employ, it is clear that not all between-college differences in completion rates are large, even if they rise to the level of statistical significance (e.g. Colleges 4 and 5). This issue, combined with the large number of "ties" between colleges, focuses attention on a problematic feature of college rankings: certain colleges are perceived as "better" than others despite relatively small, if not nonexistent, differences between them when it comes to student outcomes (Espeland and Sauder 2007).

More importantly, however, the largest problem with these rankings is that they do not distinguish between the contribution of individual students and that of the colleges they attend to producing observed outcomes. So, while graduation rates are assumed to reveal information about organizational effectiveness, they instead represent a host of factors both related and unrelated to specific colleges and universities. For example, it is well known that higher education organizations with a large proportion of high-achieving, high-income students, such as Ivy League colleges and flagship public universities, report high graduation rates (Bowen et al. 2009). It is far less clear, however, to what extent these graduation rates are related to organizational policies and practices rather than the high level of academic preparation and financial resources accompanying

the average student (Alon 2015). One could argue that the act of admitting a large proportion of highly resourced students is an organizational practice, yet it also could be viewed as the byproduct of pervasive institutional structures that systematically privilege white, higher-income students over traditionally underrepresented minority, lower-income students. In this latter scenario, numerous, non-college factors are implicated, such as students' racial and socioeconomic background, the quality of their schools prior to college entry, and the neighborhoods in which they have grown up. For this reason, I examine between- and within-college variation in effectiveness by using the ENT and WMA approaches to adequately account for students' diverse, pre-college experiences in the results reported below.

8 Average Between-College Differences in Organizational Effectiveness

I first focus on analyzing between-college variation in student outcomes employing both the ENT and WMA approaches. Table 3, Columns 2 through 5, captures the college effects coefficients for each college reported as odds ratios (Appendix Table 2 provides the full logistic regression results of both the ENT and WMA approaches using College 1 as an example). These effects represent the difference in expected BA completion between students attending the focal college as compared with those attending the average, non-focal college. However, using a "potential outcomes" approach to causality, the treatment effect of true interest is not given by the difference in the odds of BA completion between the "treated" and "untreated" groups, but rather between the treated students when pursuing their actual trajectory and *these same treated students* when pursuing a counterfactual path. While it is impossible to observe this counterfactual directly, it is useful for making empirical predictions.

I pursue this strategy here by employing predictive margins based on the initial logistic regression models, which allow me to produce expected BA completion rates for focal students in both the actual and counterfactual states. I then difference these actual and counterfactual predictions, reporting the college effect as the percentage point gain or loss in BA completion rates in Table 3, Columns 6 through 9, along with chi-square tests of these differences. Importantly, both the ENT and WMA approaches achieve extremely strong covariate balance across students

attending the focal college versus the average, non-focal college. The results reported as odds ratios and those reported as percentage point differences therefore are nearly identical, despite the fact that the comparison group is not exactly the same in the two different approaches.[11] I also find that the results generated using the alternative WMA approaches as discussed in the "Methods" section yield nearly comparable estimates (see Appendix Table 3).

Table 3 demonstrates that college effectiveness looks quite different once organizational impacts are separated from students' pre-college inputs and the selection process. College 6 now appears quite effective, increasing enrolled students' likelihood of within-college graduation by between 7 and 8 percentage points beyond students' expected completion in the average, non-focal MetroU college. In contrast, College 3 seems less effective, diminishing students' chances of within-college BA completion by between 3 and 4 percentage points. The results also demonstrate that four of the MetroU colleges, Colleges 4, 5, 7, and 9, have statistically indistinct effects on average BA completion rates, suggesting that the average student would achieve a degree at roughly the same rate if she were to attend the focal college or the average, non-focal college.

Collectively, these findings confirm my first hypothesis, that the isolation of college effects conveys different, and oftentimes contradictory, information concerning college quality as compared with aggregated graduation rates. Yet they only partially support my second hypothesis, that colleges have large, independent impacts on students' outcomes net of their background characteristics and experiences. It appears that some MetroU colleges do effect large change in the average student's outcomes, while others produce smaller or statistically indistinct impacts. This result emerges in part from the distribution of students across colleges rather than any intrinsic "quality" of each college based on the comparative principle underlying both the ENT and WMA models. Yet it still suggests that some colleges are more capable of facilitating completion among the average enrolled student than others. More technically, the results in Table 2 also convey that the findings are quite similar whether I employ the ENT or WMA approach, raising confidence in the overall robustness of the results. Given the comparability of the ENT and WMA estimates but the greater robustness and flexibility of the ENT approach, I report ENT estimates in the remainder of the paper.

[11] I also have produced the average marginal effects differences using the expected BA completion rate for students attending the average, non-focal college as the point of comparison. These results are available by request.

Table 3: Entropy Balanced and Within-Matched-Applicant Measures of College Effects on Ten-year, Within-College BA Completion Rates for Four-year Beginners

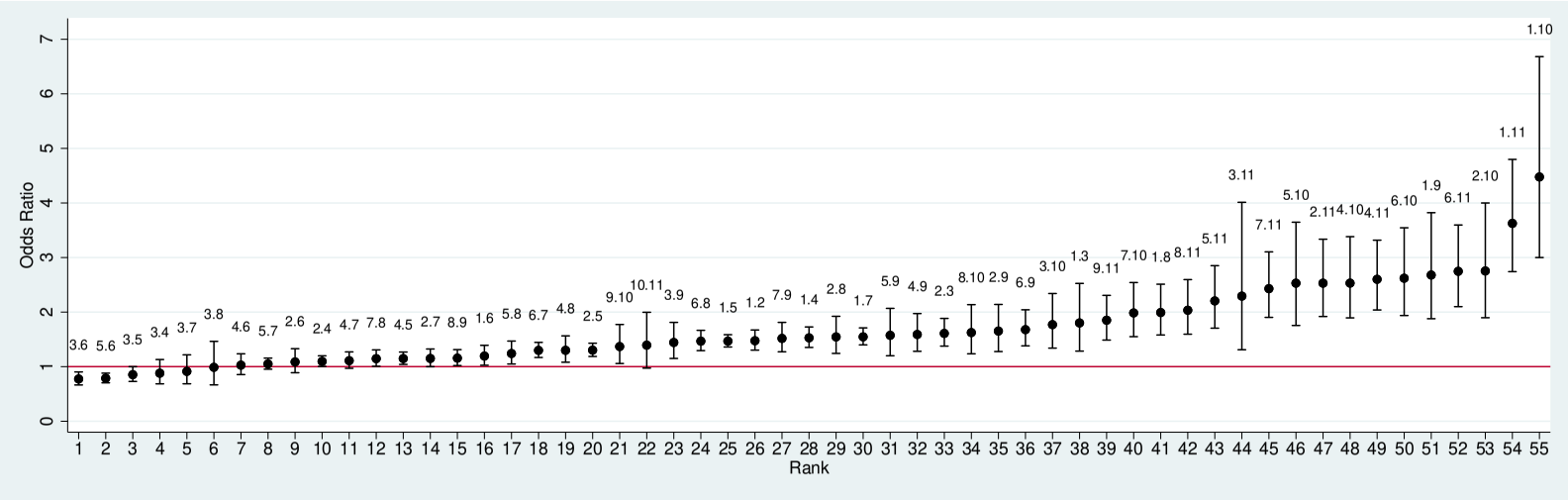
	Logistic Regression Results				Predicted Marginal Effects Differences			
	ENT Results		WMA Results		ENT Results		WMA Results	
	Odds Ratio	Standard Error	Odds Ratio	Standard Error	Dev. from Counter-factual Mean (percent-age points)	Chi-square test value	Dev. from Counter-factual Mean (percentage points)	Chi-square test value
College 1	1.251***	0.039	1.220***	0.033	4.7	0.000	4.2	0.000
College 2	0.910**	0.030	0.889***	0.029	-2.6	0.001	-2.4	0.000
College 3	0.863**	0.054	0.824***	0.051	-3.3	0.007	-4.3	0.000
College 4	0.974	0.025	0.945*	0.026	-0.6	0.280	-1.3	0.028
College 5	0.988	0.024	0.976	0.025	-0.3	0.617	0.5	0.323
College 6	1.411***	0.028	1.395***	0.029	7.7	0.000	7.4	0.000
College 7	0.957	0.025	0.960	0.026	-1.0	0.075	-0.9	0.111
College 8	0.971	0.030	0.956	0.031	-0.7	0.319	1.0	0.142
College 9	1.102	0.055	1.059	0.038	1.8	0.075	1.1	0.134
College 10	0.717***	0.063	0.738***	0.063	-6.1	0.000	-5.5	0.000
College 11	0.479***	0.106	0.528***	0.094	-12.4	0.000	-10.7	0.000

Notes: For the logistic regression coefficients, *p<0.05, **p<0.01, ***p<0.001.
Control variables in both the ENT and WMA models include student race, family income, Pell Grant status, the interaction between race and Pell Grant status, state grant aid status, special college program status, age, sex, citizenship status, dependent status, single parent status, place of permanent resident, total financial aid awarded prior to four-year college entry, whether or not a student is disabled, whether or not a student speaks proficient English, whether students attended a public or a private high school, whether the college geographically proximate to a student's high school, whether or not a student started college in the 15 months following high school graduation, high school GPA, squared high school GPA, math state test scores, English state test scores, number of college-applicable high school credits, number of MetroU colleges to which the studies applied, the year and month of college entry, and average college transfer out rates.
Source: MetroU administrative data.

It is difficult, however, to apprehend from Table 3 whether any single college is more effective than any other for the average MetroU student, since the counterfactual in these models is the average, non-focal college rather than a particular college. To address this issue, I employ the ENT approach to produce pairwise comparisons of college effectiveness, using one college within the pair as the treatment and the second as the control, eliminating all other colleges from the analysis. Figure 1 shows all pairwise comparisons in college effectiveness as odds ratios between the expected rate of BA completion for students attending the focal college versus those attending the non-focal college. They are ranked from smallest to largest and also include confidence intervals. Because they represent odds ratios, all effects that do not cross the plane at y=1 are statistically significant. In addition, because of the strong covariate balance produced by the ENT approach, these results are nearly identical to those produced using the potential-outcomes counterfactual of the BA completion rate for focal students in the average, non-focal college (see Appendix Table 4 for these alternate values).

The results in Figure 1 demonstrate that large and statistically significant differences exist in the effectiveness of some colleges over others. Of the fifty-five comparisons, forty reach the level of statistical significance. In fifteen of the pairwise comparisons, those ranked from 40 to 55, the odds of BA completion at the first college are at least two times greater than the odds at the second college. Some of these large differences are expected given the naive college rankings established in Table 2, such as between College 1 and Colleges 9, 10, and 11. However, other large differences are less expected, such as between College 7 and College 10 or College 8 and College 11, suggesting that the gap in BA completion performance between these lower-ranked colleges is more sizable than their proximate rankings suggest.

Figure 1: Pairwise Differences in College Effectiveness using Ten-year, Within-College BA Completion Rates for Four-year Beginners



Notes: College effects are estimated using the ENT approach. The labels for each estimate indicate the pairwise comparison, with the first college captured by the first term and the second found after the period. So, "3.6" represents the comparison between College 3 and College 6. Control variables include student race, family income, Pell Grant status, the interaction between race and Pell Grant status, state grant aid status, special college program status, age, sex, citizenship status, dependent status, single parent status, place of permanent resident, total financial aid awarded prior to four-year college entry, whether or not a student is disabled, whether or not a student speaks proficient English, whether students attended a public or a private high school, whether the college geographically proximate to a student's high school, whether or not a student started college in the 15 months following high school graduation, high school GPA, squared high school GPA, math state test scores, English state test scores, number of college-applicable high school credits, number of MetroU colleges to which the studies applied, the year and month of college entry, and average college transfer out rates.
Source: MetroU administrative data.

Three other notable patterns also are present. First, many colleges are no more or less effective when it comes to BA completion than those ranked immediately ahead or behind of them. For example, no statistical difference exists in the performance of College 3 as compared with Colleges 4, 5, 7 or 8 and the same can be said for College 4 as compared with Colleges 5, 6 and 7. Second, some of the non-significant distinctions are surprising in light of the naive rankings; Figure 1 shows that College 2 performs no better or worse than Colleges 6 or 7, and College 3 only performs better than Colleges 9, 10, and 11. Third, some differences that one would expect to be quite large based on the naive rankings are actually more modest. The odds ratio between Colleges 1 and 6, for example, is 1.19. Taken together, these results suggest that public perception of college quality based on unadjusted graduation rates is at times inaccurate and potentially misleading.

9 Heterogeneity in Organizational Effects

The results from Section 8 have demonstrated that sizable between-college differences exist in the average effectiveness of MetroU colleges in graduating their students, even after accounting for students' background characteristics and the selection process into college. Yet an important feature of the MetroU system, and of most public university systems, is student heterogeneity by both racial background and family income. Given this heterogeneity, it is likely that college effects are similarly heterogeneous. A college that appears to have very poor performance on average may in fact have a neutral, or perhaps even a surprisingly positive, effect on BA completion for particular student groups. Likewise, colleges with positive effects on average might in fact impact certain groups more negatively than expected. In this way, colleges may contribute more directly to the production and maintenance of inequality in students' BA completion outcomes than existing accountability metrics convey.

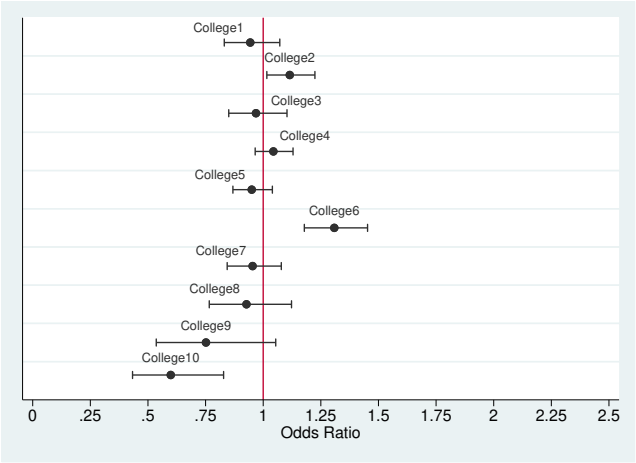
I test these interrelated hypotheses using both between- and within-college analyses, as described in the Methods section. The between-college analysis uses the same statistical formulation as in Section 8. Yet it isolates between-college, within-race/income effects by restricting the ENT models to the members of a particular racial or income group. So, for example, it shows whether College 1 is more or less beneficial than the average, other MetroU

college for the population of all black four-year beginning students. In contrast, the within-college analysis evaluates the comparative effectiveness of a single college for two, discrete student groups (e.g. white and black or non-Pell and Pell), all other variables held constant. For the within-college analyses focused on race, white students typically serve as the reference group, though I use black students as the reference for College 11 since the number of non-black students is comparably quite small. This same compositional issue for College 11 produces very large standard errors in the between-college analyses for non-black student groups, hindering graphical depiction. I therefore exclude College 11 from the graphical representations of both the between- and within-college results below, instead capturing the college effect values for College 11 (and all other colleges) numerically for the between-college analysis in Appendix Table 5 and the within-college analysis in Appendix Table 6. I report all results in this section in odds ratios, since they are more firmly rooted in the data and convey nearly equivalent trends as those produced using marginal effects models (results available by request).

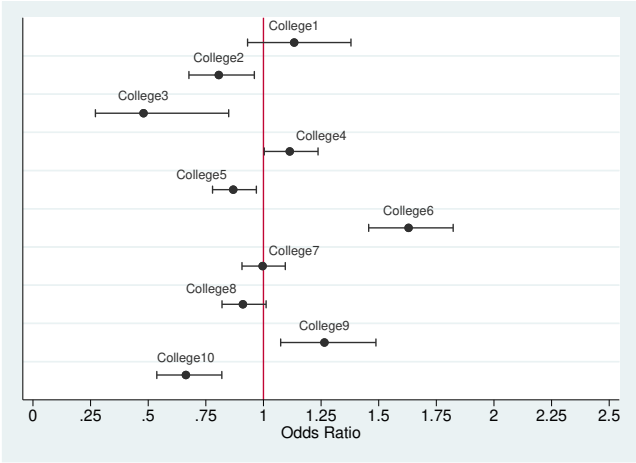
I first present the between-college results in Figure 2, which shows the impact on the odds of BA completion when attending the focal MetroU college versus the average, non-focal college for members of each, separate racial or income group. A number of surprising results emerge, in comparison with the average results reported in Table 3. First, whereas the aggregate results suggest that College 1 produces higher BA completion rates for the average student than other colleges, Figure 2 suggests that the only racial group receiving this benefit is Asian students. Students from all other racial groups are no more or less likely to collect a BA when attending College 1 than the average, other MetroU college, shown by the fact that the error bars for College 1 in subfigures A through C cross the plane at x=1. Second, though College 2 performs more poorly than expected in the aggregate, Figure 2 shows that this college in fact increases white students' odds of completion by slightly over 10 percent as compared with the other MetroU colleges. Meanwhile, College 2 also lowers the odds of completion for black, Asian, and Pell students by between 20 and 30 percent. On the other hand, College 4, which appears to lower students' odds of completion in the aggregate, in fact increases black students' odds of completion by about 10 percent but lessens the odds of completion for Hispanic and Asian students by about

Figure 2: Heterogeneity of Between-College Effects by Student Race and Income among Four-year Beginners

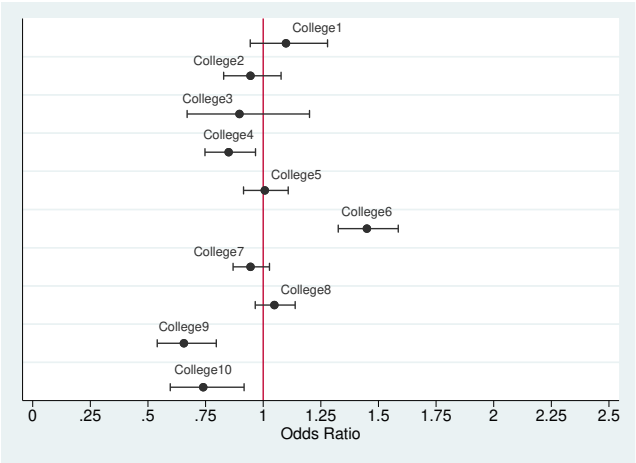
(a) White Students



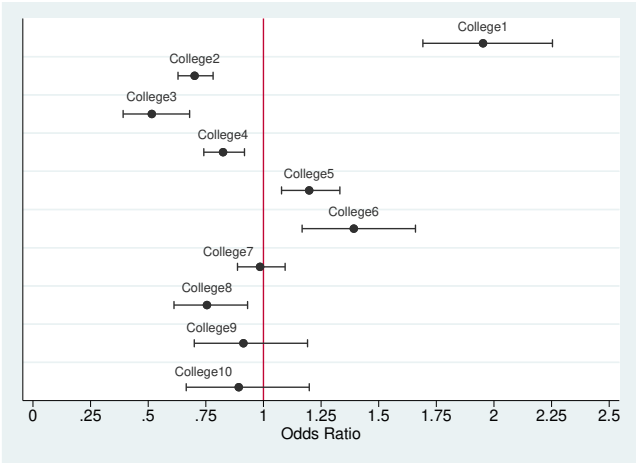
(b) Black Students



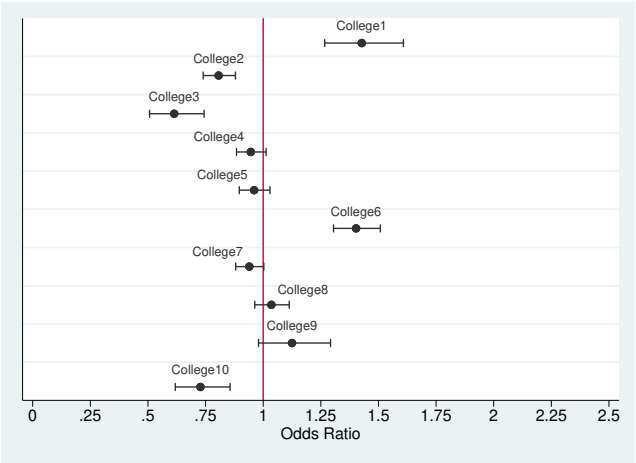
(c) Hispanic Students



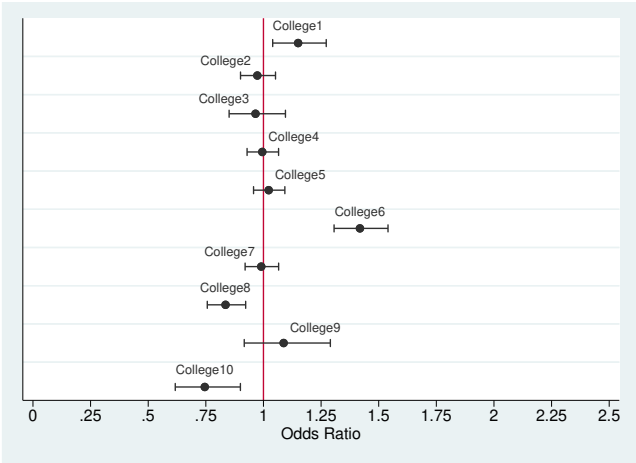
(d) Asian Students



(e) Pell Students



(f) Non-Pell Students



Source: MetroU administrative data.

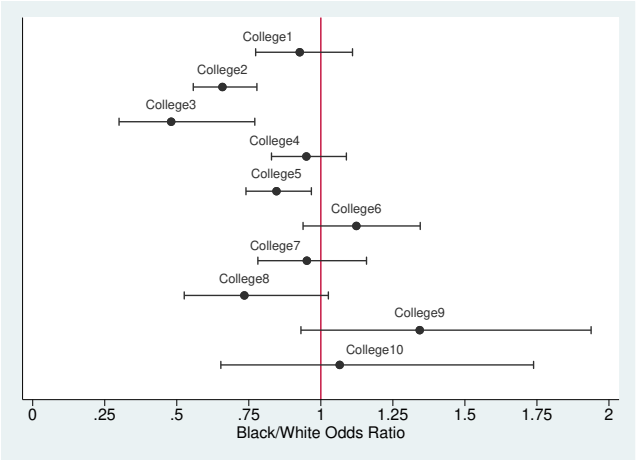
15 percent. In addition, the only MetroU college that possesses uniformly positive impacts for students, both by race and by income, is College 6, which increases the odds of BA completion for students of all racial groups and income status by between 30 and 45 percent. Considering that College 6 ranks slightly below average when comparing the colleges' unadjusted BA completion rates, these results are especially unexpected.

Though the between-college analysis provides a sense of college performance for members of each racial and income group, it cannot show the effectiveness of a particular college for members of one group versus another. For this reason, I next present the results of the within-college analysis in Figure 3, which shows the comparative odds of BA completion for racial and income dyads within ten of the eleven MetroU colleges, as described above. Focusing on the Figure 3, subfigures A through C, the results indicate that substantial within-college differences in BA completion outcomes by racial background exist for the most highly ranked colleges, in particular. Among those colleges naively ranked in the top five of the MetroU system, each possesses large inconsistencies in their graduation rates for traditionally underrepresented students as compared with white students. For example, the odds of BA completion for Hispanic students attending College 1 are 35 percent lower than for white students, while the odds for black students at Colleges 2 and 3 are 34 and 52 percent lower than for white students, respectively. Asian students do not fare much better when compared with white students, as their odds of completion are 29, 34, and 20 percent lower than white students at Colleges 2, 3, and 4. Somewhat surprisingly, the results by race appear more equitable in those colleges viewed as low performers in the naive ranking, as Colleges 6 through 10 possess little difference in the expected odds of completion for white students and other student groups. These findings in part emerge for compositional reasons: since the proportion of underrepresented minority students in quite high in these lower-ranking colleges, achieving a "balanced" comparison between white and non-white students using the ENT approach is more difficult and thus the estimates of college effects include larger standard errors.

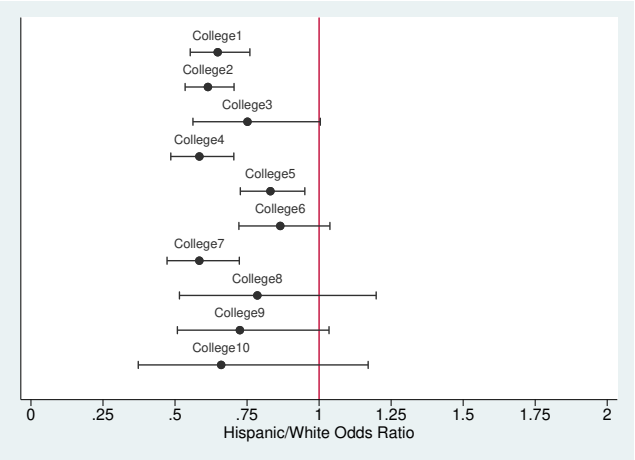
Turning to the results for non-Pell versus Pell students as reported in Figure 3, subfigure D, it appears that several of the MetroU colleges, including Colleges 1, 5, 8, 9 and 11, in fact increase the odds of BA completion for Pell students more than non-Pell students. While these effects in some ways suggest that these colleges may have an equalizing effect for Pell as compared with non-Pell students, it is also possible that Pell status may be proxying for other, important student

Figure 3: Heterogeneity of Within-College Effects by Student Race and Income among Four-year Beginners

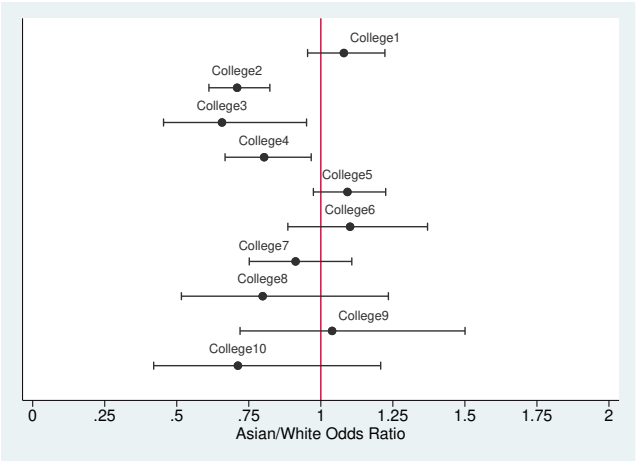
(a) Black/White Contrasts



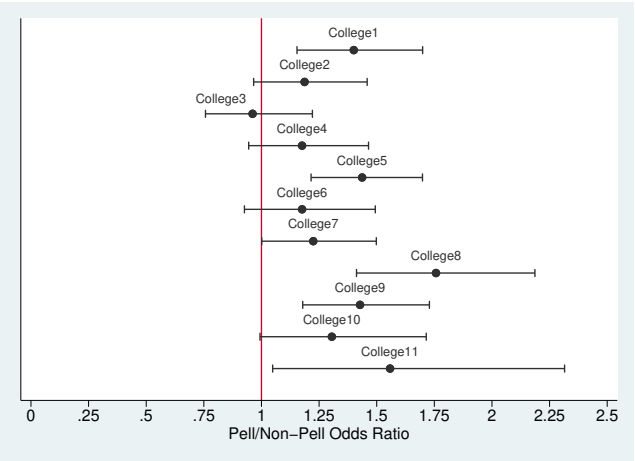
(b) Hispanic/White Contrasts



(c) Asian/White Contrasts



(d) Pell/Non-Pell Contrasts



Source: MetroU administrative data.

characteristics and experiences. This explanation may receive particular support because of the overall condensation of the family income distribution within the MetroU system, so that the difference between a student whose family earns $49,000 per year versus $51,000 per year is simply that the former student is eligible to receive federal grant support while the latter is not. Accordingly, Pell status may proxy for students' greater financial stability or greater contact with a college's administrative support services, due to the required protocol attached to receiving federal financial aid. It is also important to note that many of the confidence intervals are very wide in Figure 3, subfigure D, due to the collinearity of Pell status with other variables in the equation, especially race, as well as the large proportion of Pell students in some of the colleges, such as Colleges 8 and 11. I return to the role of college compositional characteristics in the estimation of college impacts in Section 9.

Taken together, the results examining between- and within-college heterogeneity of effects by student race and income provide strong evidence for my third hypothesis, that college effects should vary based on student characteristics. This variation occurs in two dimensions. First, certain MetroU colleges raise (or lessen) the odds of completion for students of a particular racial or income groups more so than for students of other groups. In this scenario, the point of comparison is students' odds of completion if they were to attend the average, other MetroU college. Second, most MetroU colleges lessen the odds of completion for students from at least one non-white student group as compared with white students *within the same college.* The point of comparison is a second student group rather than the average, other MetroU college in this case. Regardless of whether the between- or within-college results are evaluated, however, the findings are quite similar: publicly reported college data obscure racial and income-based inequalities in organizational effectiveness in addition to average differences in organizational effectiveness. It follows that despite the cultural commitment of American colleges and universities to equity of opportunity, that same commitment has not extended to equity of outcomes. The current accountability regime, together with the organizational dynamic of superficial coupling, enables this lack of equity, despite the goal of accountability to demonstrate greater transparency over organizational impacts on student outcomes.

Do Certain College Characteristics Explain Organizational Effectiveness?

Thus far, I have shown that between- and within-college estimates of "effectiveness" in

regards to bachelor's degree completion vary considerably in the MetroU system depending on students' pre-college characteristics, selection into college, and racial or income status. However, as mentioned at various points throughout my analysis, it is likely that certain college characteristics provide some explanation for why college effects vary as they do. College composition seems especially important within the MetroU system, as some colleges contain much higher proportions of traditionally underrepresented and lower-income students than others.

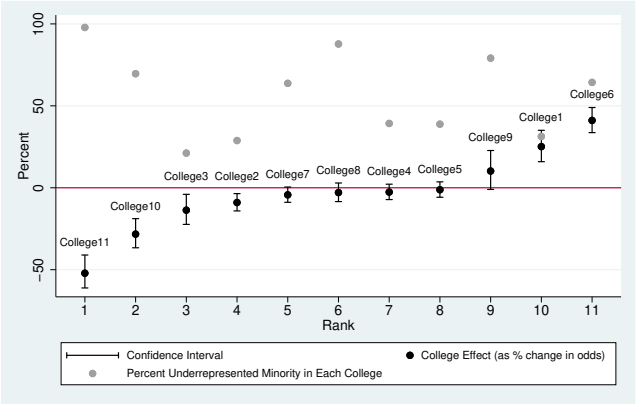
In this final set of analyses, I examine whether college-level characteristics explain any of the variation between colleges, both on average and between student groups by race and income. Specifically, I incorporate variables for college size, percent underrepresented minority, percent Pell, and a binary indicator for "mid-selective" versus "non-selective" in all estimation models, in addition to the variable for the proportion of students that transfers out. I first identify whether any of these variables predicts BA completion using mixed effects logistic regression, inserting each college into the model as a random effect. This strategy allows me to isolate the impact of college-level variables on BA completion while holding both individual-level variables and colleges constant. I begin by examining average between-college effects and then turn to between-college effects by race and income.

Figure 4, subfigures A and B, show that despite substantial compositional differences between the MetroU colleges in terms of students' racial and economic backgrounds, these discrepancies do not directly explain variation in BA completion (see Appendix Table A.7 for the regression results underlying all subfigures). . In addition, the aggregate, college-level transfer-out rate does not predict BA completion at the individual level, net of students' own transfer-out status. Instead, the sole statistically significant predictor of BA completion is college size. Here, a one standard deviation increase in college size (about 3,200 students, when combining the populations of four-year beginners and transfer students) leads to a 34 percent increase in the odds of BA completion.

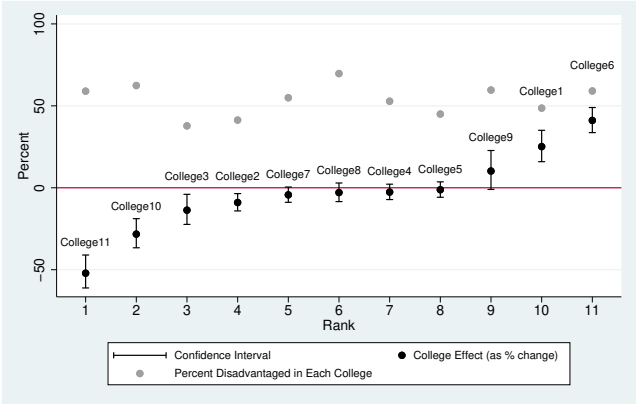
While identifying this positive relationship between college size and BA completion is generally useful, "size," itself, is a rather vague concept that likely proxies for more specific college characteristics, practices, and policies. Based on the interview research I have conducted at three of the eleven MetroU colleges (Ciocca 2017, Ciocca Eller 2019), it seems likely that some of the college qualities for which college size proxies include reputation, resource levels, and

Figure 4: The Relationship between College-level Characteristics and College Effects

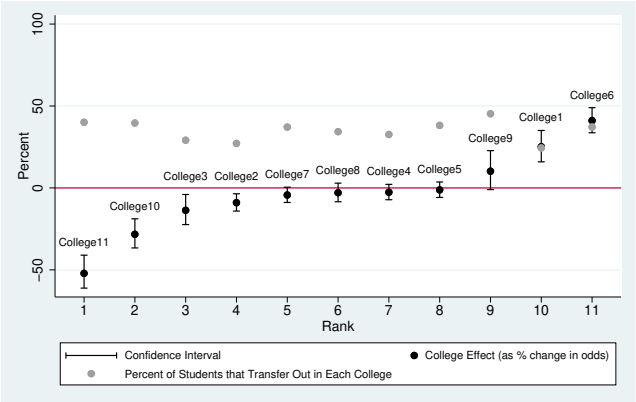
(a) Percent Underrepresented Minority



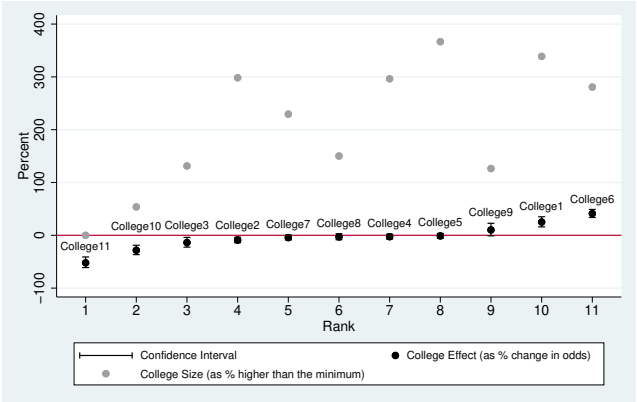
(b) Percent Pell Grant Recipients



(c) Transfer-out Rate

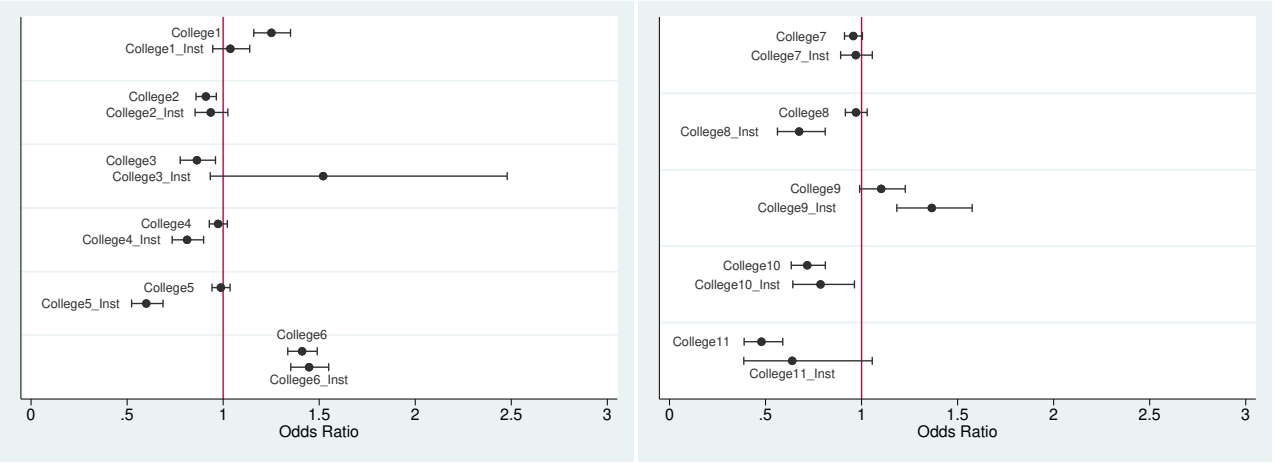


(d) College Size



Source: MetroU administrative data.

Figure 5: Comparison of Average College Effects Estimates Excluding and Including College-Level Characteristics for Four-year Beginning Students



Source: MetroU administrative data.

standardization of the college experience. In regards to reputation and resources, my interview findings suggest that the larger MetroU colleges represent those with the most public recognition and in turn, more private funding from alumni and donors. I also have found that the larger MetroU colleges possess highly structured, standardized systems for four-year beginning students, such as orientation procedures, course enrollment support, support resources to facilitate college integration, and well-publicized pathways through which students can learn about and efficiently choose a major field of study. I plan to examine the direct impacts of these college practices, as well as those of college financial resources and reputation, in future work.

The regression results just discussed show that larger colleges positively impact students' odds of BA completion. However, they do not indicate whethercollege size can explain the average, between-college variation in student outcomes identified in Section 8. I address this limitation by using the ENT method to produce additional college effects estimates that incorporate college-level characteristics. I then compare these new estimates with the effects presented in Section 8, a comparison I display graphically in Figure 5. If the points labeled with the suffix, "Inst," cross the plane at x=1 and the points with no suffix do not, then evidence exists that the transfer-out rate and college size have explained the college effect.

Figure 5 shows that these college characteristics explain some, but not all, of the between-college variation initially identified for four-year beginners. Once I equalize MetroU

colleges in regards to their size, racial and economic compositions, and selectivity rating, students' odds of BA completion in Colleges 1, 2, 3, and 11 are are no longer statistically different from their odds at the average, non-focal college. Given the regression results just discussed, part of why the average student attending College 1 can anticipate higher odds of completion than similar students enrolled in other MetroU colleges is because College 1 is larger than the other colleges. Likewise, as compared with similar students attending MetroU colleges that are larger, the students enrolled in Colleges 2, 3, and 11 can expect lower odds of completion because they are smaller. Once I control for college size, students' odds of completion within these three colleges appear comparable to the odds of students attending the average, non-focal college.

Because the MetroU system only contains eleven four-year and comprehensive colleges, it is difficult to find exact "matches" for some colleges when attempting to equalize them based on their characteristics. As discussed above, for example, College 11 contains a much higher proportion of black students than the other colleges; meanwhile College 3 is composed of a much higher proportion of white students. These compositional anomalies suggest that an additional explanation for the neutralization of college effects is simply an increase in standard error size due to high levels of statistical uncertainty. Considering the very large error bars in Figure 5 for Colleges 3 and 11, this explanation seems quite likely. That said, the regression results regarding size, as well as the relatively small error bars posted for Colleges 1 and 2, indicate that these more substantive mechanisms also are present. Future research should examine the explanatory power of transfer-out rates and college size with data on a larger set of higher education organizations, while extending this research to consider additional mechanisms. Some of these mechanisms might include the number and concentration of students within particular fields of study, the proportion of students who receive passing grades in their first semester, and the availability and quality of advising structures, among numerous others.

In a final set of analyses, I examine the explanatory power of college-level characteristics for college effects by race and income. Table 4 reports the results of separate, mixed effects regressions by race and income testing the relationship between college characteristics and BA completion. Table 4 also indicates whether the inclusion of college-level characteristics explains variation in college effects within each racial and income group by comparing the random intercept of the pre-college model (row 8), which employs the same covariates used throughout Section 9,

with that of the college model (row 9). If the random intercept value within the college model drops below the standard level of statistical significance, then the incorporation of college-level characteristics has "explained" the college effect. Notably, because the number of colleges – and hence, the statistical degrees of freedom – is quite small in these models, the elimination of statistically signification variation in the random intercept should not be interpreted as a comprehensive explanation of why certain MetroU colleges are more or less effective than others. Instead, these results should suggest that certain college characteristics incorporated into the model, such as size or racial / economic composition, are strongly correlated with college effectiveness for certain student groups.

The results convey that the addition of college-level covariates provides more explanatory power for between-college variation in BA completion for white and Hispanic students as compared with black and Asian students. Among white students, a college's transfer-out rate has the most explanatory power, with a one standard deviation increase in the transfer-out rate decreasing white students' odds of completion by about 20 percent, all other covariates held constant. Since the covariates in these explanatory models include transfer at the individual level, the negative effect of college-level transfer-out rates on BA completion for white students occur net of individual students' own transfer trajectories. Once white students are equalized based on their colleges' transfer-out rate and the additional college-level variables, they appear no more or less likely to earn a BA from any, single MetroU college. Turning to Hispanic students, college size is the strongest predictor of college effectiveness, as it is among the overall population of MetroU students. Here, a one standard deviation increase in college size results in a 38 percent increase in Hispanic students' odds of completion, net of other covariates. As with white students, the inclusion of college size and the other college-level variables erases the statistical presence of between-college variation in the odds of BA completion for Hispanic students.

Meanwhile, despite the fact that certain college-level variables appear predictive of BA completion for students of other racial and income groups, their addition to the model does not fully explain between-college variation in the odds of degree attainment for black, Asian, Pell and non-Pell students. In considering why these differences in the explanatory power of college-level variables by race and income occur, several issues emerge.

Table 4: The Effects of College-level Characteristics on BA Completion for Four-year Beginners by Student Race and Income, as Odds Ratios

	White (*n=26,606*)		Black (*n=17,428*)		Hispanic (*n=20,490*)		Asian (*n=17,461*)		Pell (*n=40,518*)		Non-Pell (*n=41,467*)	
	Coef.	*Std. Err.*	*Coef.*	*Std. Err.*	*Coef.*	*Std. Err.*	*Coef.*	*Std. Err.*	*Coef.*	*Std. Err.*	*Coef.*	*Std. Err.*
Transfer-out Rate (std)	0.794***	0.048	0.957	0.058	0.891	0.085	0.841	0.202	0.870*	0.051	0.917	0.059
Size (std)	1.132	0.136	1.256	0.160	1.384*	0.217	1.364*	0.202	1.396**	0.166	1.224	0.159
Percent Underrepresented Minority (std)	1.118	0.195	0.917	0.144	0.841	0.247	1.042	0.284	1.013	0.153	0.807	0.134
Percent Pell Grant recipient (std)	0.919	0.115	1.251*	0.142	1.221	0.255	1.083	0.212	1.103	0.122	1.159	0.141
Mid-Selective rating	1.151	0.231	1.136	0.245	0.773	0.212	1.008	0.253	0.826	0.165	0.924	0.201
Random Intercept of pre-college model (std. dev.)	0.207***	0.062	0.319***	0.074	0.329***	0.112	0.289***	0.079	0.280***	0.065	0.325***	0.077
Random Intercept of college model (std. dev.)	0.085	0.044	0.124***	0.033	0.160	0.061	0.140***	0.041	0.122***	0.030	0.132***	
Individual-level controls	X		X		X		X		X		X	
Month, year, and application controls	X		X		X		X		X		X	
Constant	0.189***	0.050	0.177***	0.045	0.382**	0.128	0.553	0.202	0.188	0.235	0.304***	0.067

Notes: The omitted category for "mid-selective rating" is non-selective rating.
*p<0.05, **p<0.01, ***p<0.001.
Source: MetroU administrative data.

First, it is perhaps not surprising that college-level variables do not provide substantially more explanatory power for between-college variation among either Pell or non-Pell students based on a) the condensed income distribution among MetroU students; and b) the inability to explain

average between-college variation in BA completion through the inclusion of college-level characteristics. Since Pell status splits the MetroU student distribution roughly in half, the results recorded in Table 4, columns 10-13, are quite expected.

Second, the initial level of between-college variation for white students is lower than that of any other racial or income group. As a result, the explanatory power of college-level variables among white students need not be as great as for the other student groups in order to diminish the statistical significance of the random intercept. In other words, while college-level variables potentially are substantively important in explaining between-college variation in BA completion rates for white students, it also is likely that the initially lower level of between-college variation is reflected in the lessening statistical significance of the random intercept for this group. Translating this logic to the case of Hispanic students, the high level of initial variation in college effectiveness for this group makes it more likely that college-level variables provide a substantively meaningful explanation for between-college effects.

Beyond the importance of college size, and the processes for which it proxies, my interview data suggest that Hispanic students are more sensitive and responsive to the racial and ethnic composition of their college than are other student groups. Specifically, Hispanic students appear to rely more often on students from common backgrounds than do students from other racial groups for support and encouragement, a dynamic also documented in existing research (Charles et al. 2007). While the significance of college composition for Hispanic students is not represented statistically in Table 4, it is possible that the effects flow through, and in interaction with, other college-level variables in the model, such as size and the transfer-out rate. Future research should investigate these possibilities.

10 Discussion

In this paper, I have proposed that higher education organizations maintain public legitimacy and ensure survival without demonstrating their unique role in producing student outcomes or fulfilling their culturally sanctioned purpose of creating student opportunity. They are able to do so because of deep and enduring tensions between the institutional logics, or "rationalized myths" (Meyer and Rowan 1977), of accountability and American individualism. Whereas the logic of

accountability has facilitated the rise of governmental policies and normative cultural practices that require college and university leaders to report data on college outcomes, the logic of American individualism transfers the responsibility for success or failure onto the shoulders of students. The data typically required for public reporting purposes, average bachelor's degree completion rates, encapsulate this dynamic, as they technically depict BA completion at the individual level rather than demonstrating the role of colleges and universities in producing those BA completion rates. As a result, numerous, important trends in student outcomes are masked, including average differences in BA completion rates between colleges and universities with similar levels of selectivity, as well as inequality in between- and within-college inequality in BA completion by student race and income. At the organizational level, this process has translated to a new structural arrangement, "superficially coupled systems," in which college and university leaders closely oversee the production of technical outputs (e.g. BA completion rates) while implicitly shielding their organizations from direct implication in the production of often unequal outputs. In this sense, the apparent tightening between organizational oversight and technical outputs demanded by the logic of accountability is only topical in its effects, useful for maintaining organizational legitimacy but not for improving student outcomes or shedding light on enduring inequalities between student groups.

My findings provide empirical evidence to support this argument. Once I employ statistical measures that isolate the role of colleges in the production of student BA completion outcomes, some colleges that appear quite successful by virtue of their naive, average BA completion rates in fact are shown to produce limited, or even negative, impacts on the average student's likelihood of completion. For example, students attending the colleges ranked second and third in the MetroU system by way of BA completion rates actually possess odds of BA completion 10 and 15 percent lower, respectively, than similar students attending the average, non-focal college. Meanwhile, students enrolled in the college ranked sixth by its BA completion rate can anticipate odds of completion 40 percent higher than similar students attending the average, non-focal MetroU college. These findings suggest that current rankings based on publicly available data often are inaccurate and sometimes are misleading.

The results also demonstrate between- and within-college heterogeneity in effectiveness by student background, though more by race than by family income. The between-college results

suggest that high levels of college effectiveness in producing average completion rates rarely translates to equally beneficial effects for all racial or income groups. Evidence from College 1 is a good example, as the aggregated results suggests a 25 percent increase in the odds of completion for the "average" student, while the results by race show that only Asian students experience positive, significant odds of completion when attending College 1 while all other groups experience just average outcomes. In fact, only one college, College 6, uniformly lifts students' odds of BA completion across all racial and income groups. These findings align with research developed in the secondary education context suggesting that "effectiveness" and "equity" are distinct axes along which schools (or colleges) can, and should, be evaluated (Lee et al. 1997).

The within-college results confirm and add nuance to these findings by testing college effectiveness in pairwise comparisons between dominant and non-dominant student groups. They indicate that in nearly half of the MetroU colleges, at least one, non-white group possesses lower odds of BA completion than white students. While some colleges appear to lift the odds of BA completion more for low-income students (e.g. Pell recipients) more than higher-income students (e.g. non-Pell recipients), it is possible that issues besides college effectiveness underlie these findings, as discussed above. Furthermore, in examining the the colleges where inequality by student race and income is greatest, nearly all are naively ranked in the top four of the MetroU system. The inconsistency between these colleges' top rankings and their production of inequality further solidifies the importance of rethinking our methods for establishing the "best" colleges. Specifically, should colleges with the highest graduation rates be viewed as "the best" if they also contribute to tremendous inequality between student groups? Considering the role policy-makers and the general public expect colleges to play in ensuring equality of opportunity, the maintenance of ratings systems and data collection efforts that privilege absolute over relative success seems contrary to this widespread expectation.

College-level characteristics can explain some of the variation in college effectiveness. The fact that college size matters for BA completion for four-year beginning students suggests that colleges have the opportunity to address poor or unequal BA completion outcomes by recognizing organizational characteristics, policies, and practices that directly impact their students. In addition, the fact that college characteristics can explain between-college differences in BA completion outcomes for certain student groups, such as white and Hispanic students, but not for

others, such as black and Asian students, further heightens the importance of interrogating college qualities and procedures to ameliorate student outcomes. Certainly, it is difficult to change longstanding college characteristics such as size. Yet as discussed above, college size likely proxies for more substantial policies and practices that could elevate and/or equalize students' chances of earning a degree. Two of these include financial support and structure of the academic and social curriculum. More in-depth observational and interview-based research would help to identify additional mechanisms related to college size while also helping to identify other college-level characteristics and practices that shape student outcomes. Given Coleman et al.'s (1966) early work, it is likely that peer effects and teacher effects, especially in regards to academic and emotional guidance and support, play important roles. I explore these issues in related research (Ciocca Eller 2019).

Despite the importance of its findings, this paper only can conjecture, rather than conclusively prove, whether the structural dynamics of superficially coupled systems occur purposefully or as a habituated response to the current system of higher education accountability. This limitation is similar to that encountered by Meyer and Rowan (1977, 1978), whose work provides minimal empirical evidence showing that bureaucratic leaders *intentionally* distance themselves from the technical processes and outputs of their organizations as a *strategy* for maintaining social legitimacy. In the present context, it is clear that the leaders of public higher education organizations have strong financial and reputational incentives to report student outcomes that are aligned with performance goals – especially in an era of performance-based funding. It is also clear that one lever that higher education leaders have used to achieve such outcomes (recall, these mainly are related to average BA completion rates) is to adjust the student body composition through the admissions process. Specifically, scholars have shown that one of the "unintended consequences" of performance-based funding in higher education is to produce more restrictive college admissions policies that privilege higher-income, less racially diverse, and higher-achieving students (Dougherty et al. 2016a; Kelchen and Stedrak 2016). Such a shift at the point of admissions is potentially easier for college and university administrators to achieve than a vast overhaul of college practices and policies, which might be required if performance metrics shifted to isolate organizational impacts. In short, just as there are incentives for the leaders of public higher education organizations to meet performance goals, there also are more nuanced,

and less explicit, incentives for these leaders to conform to the existing system of accountability rather than to reshape it in ways that make organizational impacts clearer and more actionable. While additional data are required to verify this argument, perhaps including interviews with and observations of high-level university leaders and institutional research offices, at least some intentionality appears likely in the production of superficially coupled systems.

In regards to other issues this study has not addressed, the question of whether "BA completion" means the same at each of MetroU's four-year and comprehensive colleges remains. This issue also raises the additional point of whether "organizational effectiveness" is a consistent concept across colleges. It may be possible, for example, that colleges are more or less effective in enabling BA completion due to easier or more difficult completion requirements, respectively, rather than because of any particular organizational action (besides setting degree completion requirements). As recent controversies in urban higher education systems have shown, such as those within the two-year public system in Chicago (Smith 2017), this type of "gaming" of the system can occur. Such gaming also would align with arguments suggesting that student learning is relatively minimal in four-year colleges and universities (Arum and Roksa 2011), inflating the number of BAs granted while lessening the value of the credential in the labor market. It also could be the case that students perceive BAs from certain MetroU colleges as more valuable than BAs from others, motivating students differently to complete degrees based on the college they attend. This circumstance again would inject bias into results quantifying college effectiveness, as students' perceptions of value and related actions would serve as drivers of organizational effects rather than the organization, itself.

In the present context, however, both of these scenarios are unlikely. In the first instance, the fact that all eleven MetroU colleges are part of a single university system means that each of the colleges must comply with a system-wide definition of acceptable degree completion standards. Accordingly, it is reasonable to imagine that BA completion and college effectiveness are comparable concepts across colleges in the MetroU system and that "gaming" either is taking place uniformly or not at all. In the second instance, though it is not possible to account directly for student perceptions concerning the value of a BA at various of the MetroU colleges, their initial application preferences are a strong and compelling proxy. Because these application preferences are available in the MetroU data, and accounted for in the statistical models used throughout this

paper, it is unlikely that students' differing perceptions of the meaning and value of a BA across the MetroU colleges has been conflated with the identified college effects.

11 A Framework for Future Research

Beyond these data-driven and technical issues, this study raises a number of substantive questions that require additional consideration. First, I have argued throughout the paper that average BA completion rates are inadequate indicators of college effectiveness. Yet I have not addressed the important, related question of how to interpret adjusted college effects estimates in conjunction with unadjusted graduation rates. For example, although colleges ranked in the lower half of the MetroU distribution tend to produce more equitable results for student groups, they still possess BA completion rates hovering between 30 and 50 percent ten years after college entry, even when using the most lenient definition of BA completion, or degree receipt at any college. It is unclear if equity should be celebrated in this context, considering the relatively limited ability for enrolled students to earn their BA. In a best-case scenario, higher education organizations would elevate the overall level of student success so that trade-offs between equity and excellence would not be necessary. In the meantime, scholars and policy-makers should direct more concerted thought to balancing equity and excellence when assessing college performance.

Second, this study also has assumed that the most important outputs of higher education organizations are bachelor's degrees. In doing so, it implicitly, though unintentionally, has supported Meyer and Rowan's (1978) argument that the purpose of colleges and universities predominantly is to provide credentials rather than to facilitate students' academic learning, socio-emotional growth, or development as citizens (Arum and Roksa 2011; Binder and Wood 2013; Guhin 2016; Labaree 1997). It also has not considered the broader roles of colleges and universities as sites of scholarly research, open discourse and exchange, and the construction of important relationships, both at the individual or university level (Cole 2011; Stevens and Gebre-Medhin 2016; Stevens et al. 2008). Part of why contemporary social science scholarship focuses less on these important features of the American university system is the limited availability of high-quality data linking non-curricular experiences and outcomes with individual colleges. Some research has moved in this direction (e.g. Armstrong and Hamilton 2013; Bowen

and Bok 2000; Chambliss and Takacs 2014; Charles et al. 2009; Massey et al. 2011; Stuber 2012), yet most of these existing studies focus on single cases, comparisons between two colleges, or on elite higher education organizations. Considering the heightened discrepancies between elite and non-elite colleges in the types of academic and social experiences available to students (Brint et al. 2012; Rosenbaum et al. 2006; Kirst and Stevens 2015), future research should continue to study how distinct college characteristics and practices divide students using dependent variables other than grade-point average and BA completion.

Another reason why less research is produced regarding non-curricular outcomes, however, is because of the enduring emphasis in academic and political circles on concrete, measurable outcomes. As with graduation rate data, this emphasis has emerged alongside the rise of accountability standards and its accompanying belief in quantification. However, this intensive attention to the measurement of outcomes ignores two important facts: (1) not all important outcomes are easily measurable, and (2) decisions concerning which educational outcomes deserve quantification, and how that quantification occurs, are highly subjective. Regarding the first point, scholars have argued that the outputs of bureaucratic organizations are difficult to identify and measure because the primary purpose of bureaucracies is not to produce any particular output (e.g. Wilson 1991). Instead, bureaucracies are "coping organizations," in which most time and attention of personnel are directed towards "coping" with organizational norms, rules, and constraints rather than efficiently creating a product. Accordingly, many scholars of organizations have employed ethnographic approaches in an effort to understand the structure and outcomes of greatest importance in these settings (recently, for example, Lara-Millán 2014; Vaughan and Vaughan 1996; Watkins-Hayes 2009). Yet given the centrality of non-technical outcomes to individuals' organizational experiences, large-scale data collection efforts that would enable statistical evaluation of these alternative outcomes are an important future step in the study of organizational impacts on individual opportunity.

Regarding the second point, this paper has shown that the production of different kinds of statistical estimates leads to the production of different kinds of knowledge. One explanation for the normative reliance on aggregated graduation rates is practicality, as more robust statistical calculations are labor-intensive and thus inefficient from an organizational standpoint. Another is insufficient awareness: the use of data analytics in higher education is a relatively new practice,

and perhaps is overshadowed by a combination of inscribed organizational practices, which endure due to inertia, and the numerous "fires" college and university administrators must put out on a daily basis. However, several large university systems in the United States, such as those in Texas and Ohio, have begun to use data analysis more extensively as a means of ameliorating gaps in college outcomes between traditionally underrepresented minority students and non-underrepresented students (Alvarado et al. 2018; Nichols et al. 2016). This evolution represents an important development with the potential to press public colleges and universities to acknowledge their direct organizational influence on student outcomes. Of course, it is possible that increased reliance on more statistically sophisticated analytics instead could serve as another strategy for fulfilling the legitimacy imperatives of higher education organizations in formal but not substantive terms. In other words, such analytics may provide college and university leaders with stronger evidence to suggest they have "done all they can" to facilitate positive student outcomes without taking the necessary actions to produce those outcomes, whether in relation to BA completion or other important outcomes like labor market placement or student learning (Arum and Roksa 2016). In this sense, even statistically robust evidence of college impact is insufficient if embedded within an institutional logic of accountability that does not, in fact, provide an adequate framework for holding organizations accountable.

Third, though the processes described throughout this work apply to the American higher education system broadly construed, the status of Metropolitan University as a public system entails higher stakes when it comes to navigating the tension between accountability and American individualism. Accountability pressures are intensified for public colleges and universities because of their reliance on taxpayer dollars for financial solvency. At the same time, public colleges and universities serve nearly 75 percent of all college-goers, a large proportion of whom come from traditionally underrepresented minority and low-income backgrounds (Snyder et al. 2018). Public colleges and universities therefore must demonstrate compliance with accountability reporting standards on the one hand while addressing substantial challenges, such as decreased funding at the state and federal levels and often-insufficient academic preparation among individual students due to uneven elementary and secondary schooling, on the other. From a broader, cultural perspective, public higher education organizations also must prioritize their moral role as the bearers of opportunity, knowledge, and upward social mobility for their students

in order to maintain public confidence and continued legitimacy. The complexities of managing these priorities at the organizational level are extensive, likely contributing to the dynamic of superficially coupled systems. Yet without an accountability regime that both sheds light on the direct contribution of colleges and universities to student outcomes, and then requires colleges and university leaders to act on those findings, it is unlikely that higher education organizations will discover strategies for addressing their continued legitimacy and survival while also creating better and more equitable student outcomes.

Finally, while the present paper takes higher education as its empirical case, it is likely that the processes underlying superficially coupled systems also extend to other complex bureaucratic systems, such as health care or the federal government. In these additional sectors, the presence of institutionalized legitimacy standards, combined with entrenched organizational structures, practices, and interests, likely conceals inequality in individual patient health or the individual receipt of federal aid dollars, to name just two possible outcomes. A number of scholars have begun to trace these processes (e.g. Reich 2014; Watkins-Hayes 2009), though not using the logical framework of superficially coupled systems. Considering the continued centrality of accountability as a governing logic within numerous social sectors in the United States, understanding whether existing accountability regimes, in combination with organizational responses to those regimes, perpetuate inequality is particularly important work.

References

Alexander, Karl and Stephen L. Morgan. 2016. "The Coleman Report at Fifty: Its Legacy and Implications for Future Research on Equality of Opportunity." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 2(5):1–16.

Allison, Paul. 2002. *Missing Data*. Thousand Oaks, CA: SAGE Publications, Inc.

Alon, Sigal. 2015. *Race, Class, and Affirmative Action*. New York, NY: Russell Sage Foundation.

Alon, Sigal and Marta Tienda. 2005. "Assessing the "Mismatch" Hypothesis: Differences in College Graduation Rates by Institutional Selectivity." *Sociology of Education* 78(4):294–315.

Alvarado, Cassandre Giguere, Carolyn Connerat, and Jennifer L. Smith. 2018. "Every Student Graduates: The University of Texas at Austin Strategies to Increase Graduation, 2012-2017." Technical report, University of Texas at Austin, Austin, TX.

Angrist, Joshua D. and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, N.J.: Princeton University Press.

Armstrong, Elizabeth A. and Laura T. Hamilton. 2013. *Paying for the Party: How College Maintains Inequality*. Cambridge, MA: Harvard University Press.

Arum, Richard and Josipa Roksa. 2011. *Academically Adrift: Limited Learning on College Campuses*. Chicago, IL: University of Chicago Press.

Arum, Richard and Josipa Roksa. 2016. "Measuring College Performance." In *Remaking College: The Changing Ecology of Higher Education*, edited by Michael W. Kirst and Mitchell L. Stevens, pp. 169–189. Stanford University Press.

Bailey, Thomas R., Shanna Smith Jaggers, and Davis Jenkins. 2015. *Redesigning America's Community Colleges*. Harvard University Press.

Berger, Peter L. and Thomas Luckmann. 1967. *The Social Construction of Reality: A Treatise in the Sociology of Knowledge*. New York, NY: Penguin Books.

Binder, Amy. 2007. "For Love and Money: Organizations' Creative Responses to Multiple Environmental Logics." *Theory and Society* 36(6):547–571.

Bourdieu, Pierre. 1977. *Outline of a Theory of Practice*. Cambridge, UK: Cambridge University Press.

Bourdieu, Pierre and Jean-Claude Passeron. 1979. *The Inheritors: French Students and Their Relation to Culture*. Chicago, IL: The University of Chicago Press.

Bowen, William G. and Derek Bok. 2000. *The Shape of the River: Long-Term Consequences of Considering Race in College and University Admissions*. Princeton, N.J.: Princeton University Press.

Bowen, William G, Matthew M Chingos, and Michael S McPherson. 2009. *Crossing the finish line: completing college at America's public universities*. Princeton, N.J.: Princeton University Press.

Brand, Jennie E. and Charles N. Halaby. 2006. "Regression and matching estimates of the effects of elite college attendance on educational and career achievement." *Social Science Research* 35(3):749–770.

Brint, Steven and Charles T. Clotfelter. 2016. "U.S. Higher Education Effectiveness." *RSF* 2(1):2–37.

Brint, Steven, Kristopher Proctor, Kerry Mulligan, Matthew B. Rotondi, and Robert A. Hanneman. 2012. "Declining Academic Fields in U.S. Four-Year Colleges and Universities, 1970-2006." *The Journal of Higher Education* 83(4):582–613.

Carnevale, Anthony P. 2008. "College For all?" *Change: The Magazine of Higher Learning* 40(1):22–31.

Chambliss, Daniel F. and Christopher G. Takacs. 2014. *How College Works*. Harvard University Press.

Charles, Camille Z., Mary J. Fischer, Margarita A. Mooney, and Douglas S. Massey. 2009. *Taming the River: Negotiating the Academic, Financial, and Social Currents in Selective Colleges and Universities*. Princeton University Press.

Chetty, Raj, John Friedman, Emmanuel Saez, Nicholas Turner, and Danny Yagan. 2017. "Mobility Report Cards: The Role of Colleges in Intergenerational Mobility." Technical Report w23618, National Bureau of Economic Research, Cambridge, MA.

Ciocca Eller, Christina. 2019. "Superficially Coupled Systems: The Organizational Production of Educational Inequality." Unpublished dissertation paper.

Ciocca Eller, Christina and Thomas A. DiPrete. 2018. "The Paradox of Persistence: Explaining the Black-White Gap in Bachelor's Degree Completion." *American Sociological Review* 83(6):1171–1214.

Cole, Jonathan R. 2011. *The Great American University: Its Rise to Preeminence, Its Indispensable National Role, Why It Must Be Preserved*. New York, NY: Public Affairs.

Coleman, James S., Ernest Q. Campbell, Carol J. Hobson, James McPartland, Alexander M. Mood, Frederic D. Weinfeld, and Robert L. York. 1966. "Equality of Educational Opportunity." Technical report, Government Printing Office, Washington, DC.

Collins, James. 1988. "Language and Class in Minority Education." *Anthropology & Education Quarterly* 19(4):299–326.

Cowan, Jon and Jim Kessler. 2017. "Opinion | How to Hold Colleges Accountable." *The New York Times* .

Crosnoe, Robert, Monica Kirkpatrick Johnson, and Glen H. Elder. 2004. "Intergenerational Bonding in School: The Behavioral and Contextual Correlates of Student-Teacher Relationships." *Sociology of Education* 77(1):60–81.

Cunha, Jesse M. and Darwin W. Miller. 2009. "Quantitative Measures Of Achievement Gains And Value-Added In Higher Education: Possibilities And Limitations In The State Of Texas." Technical report, Texas Higher Education Coordinating Board.

Cunha, Jesse M. and Darwin W. Miller. 2014. "Measuring value-added in higher education: Possibilities and limitations in the use of administrative data." *Economics of Education Review* 42:64–77.

Dale, Stacy Berg and Alan B. Krueger. 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, Stacy B. and Alan B. Krueger. 2014. "Estimating the Effects of College Characteristics over the Career Using Administrative Earnings Data." *Journal of Human Resources* 49(2):323–358.

Davies, Mark and Denise B. Kandel. 1981. "Parental and Peer Influences on Adolescents' Educational Plans: Some Further Evidence." *American Journal of Sociology* 87(2):363–387.

DiMaggio, Paul J. and Walter W. Powell. 1983. "The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields." *American Sociological Review* 48(2):147–160.

DiMaggio, Paul J. and Walter W. Powell. 1991. "Introduction." In *The New Institutionalism in Organizational Analysis*, edited by Walter W. Powell and Paul J. DiMaggio, pp. 1–40. Chicago, IL: The University of Chicago Press.

Dougherty, Kevin J., Sosanya M. Jones, Hana Lahr, Rebecca S. Natow, Lara Pheatt, and Vikash Reddy. 2016a. "Looking Inside the Black Box of Performance Funding for Higher Education: Policy Instruments, Organizational Obstacles, and Intended and Unintended Impacts." *RSF* 2(1):147–173.

Dougherty, Kevin J., Sosanya M. Jones, Hana Lahr, Lara Pheatt, Rebecca S. Natow, and Vikash Reddy. 2016b. *Performance Funding for Higher Education*. JHU Press.

Downey, Douglas B. and Dennis J. Condron. 2016. "Fifty Years since the Coleman Report: Rethinking the Relationship between Schools and Inequality." *Sociology of Education* 89(3):207–220.

Downey, Douglas B. and Shana Pribesh. 2004. "When Race Matters: Teachers' Evaluations of Students' Classroom Behavior." *Sociology of Education* 77(4):267–282.

Espeland, Wendy Nelson and Michael Sauder. 2007. "Rankings and Reactivity: How Public Measures Recreate Social Worlds." *American Journal of Sociology* 113(1):1–40.

Espeland, Wendy Nelson and Michael Sauder. 2016. *Engines of Anxiety: Academic Rankings, Reputation, and Accountability*. New York, NY: Russell Sage Foundation.

European Commission. 2015. "Dropout and Completion in Higher Education in Europe." Technical report, European Union, Luxembourg.

Ewell, Peter T. 2010. "Student-Unit Record Systems and Postsecondary Accountability: Exploiting Emerging Data Resources." In *Accountability in American Higher Education*, edited by Kevin Carey and Mark Schneider. New York, NY: Palgrave McMillan.

Fink, John and Davis Jenkins. 2017. "Takes Two to Tango: Essential Practices of Highly Effective Transfer Partnerships." *Community College Review* 45(4):294–310.

Fischer, Mary J. and Douglas S. Massey. 2007. "The effects of affirmative action in higher education." *Social Science Research* 36(2):531–549.

Friedland, Roger and Robert R. Alford. 1991. "Bringing Society Back In: Symbols, Practices, and Institutional Contradictions." In *The New Institutionalism in Organizational Analysis*, edited by Walter W. Powell and Paul J. DiMaggio, pp. 232–266. Chicago, IL: The University of Chicago Press.

Gamoran, Adam and Robert D. Mare. 1989. "Secondary School Tracking and Educational Inequality: Compensation, Reinforcement, or Neutrality?" *American Journal of Sociology* 94(5):1146–1183.

Gándara, Denisa and Amanda Rutherford. 2018. "Mitigating Unintended Impacts? The Effects of Premiums for Underserved Populations in Performance-Funding Policies for Higher Education." *Research in Higher Education* 59(6):681–703.

Garfinkel, Harold. 1967. *Studies in Ethnomethodology*. Englewood Cliffs, NJ: Prentice-Hall.

Gerber, Theodore P. and Sin Yi Cheung. 2008. "Horizontal Stratification in Postsecondary Education: Forms, Explanations, and Implications." *Annual Review of Sociology* 34(1):299–318.

Goffman, Erving. 1967. *Interaction Ritual: Essays in Face to Face Behavior*. Chicago, IL: Aldine Publishing Company.

Goodman, Joshua, Michael Hurwitz, and Jonathan Smith. 2015. "College Access, Initial College Choice and Degree Completion." Working Paper 20996, National Bureau of Economic Research.

Greenwood, Royston, Christine Oliver, Kerstin Sahlin, and Roy Suddaby. 2008. "Introduction." In *The SAGE Handbook of Organizational Institutionalism*, pp. 1–46. London: SAGE Publications Ltd.

Grodsky, Eric. 2007. "Compensatory Sponsorship in Higher Education." *American Journal of Sociology* 112(6):1662–1712.

Grubb, W. Norton and Marvin Lazerson. 2004. *The Education Gospel: The Economic Power of Schooling*. Cambridge, MA: Harvard University Press.

Guhin, Jeffrey. 2016. "Why sociologists of education are moral (personally) but don't write about morality (professionally): a cultural sociology of education." *Sociologia e Politiche Sociali* pp. 34–51.

Hainmueller, Jens. 2012. "Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies." *Political Analysis* 20(01):25–46.

Hainmueller, Jens and Yiqing Xu. 2013. "ebalance: A Stata Package for Entropy Balancing." *Journal of Statistical Software* 54(7):1–18.

Hallett, Tim. 2010. "The Myth Incarnate: Recoupling Processes, Turmoil, and Inhabited Institutions in an Urban Elementary School." *American Sociological Review* 75(1):52–74.

Hallett, Tim and Marc J. Ventresca. 2006. "How Institutions Form: Loose Coupling as Mechanism in Gouldner's Patterns of Industrial Bureaucracy." *The American Behavioral Scientist* 49(7):908–924.

Hallinan, Maureen T. and Richard A. Williams. 1990. "Students' Characteristics and the Peer-Influence Process." *Sociology of Education* 63(2):122–132.

Hanushek, Eric A. 1997. "Assessing the Effects of School Resources on Student Performance: An Update." *Educational Evaluation and Policy Analysis* 19(2):141–164.

Heck, Ronald H. 2000. "Examining the Impact of School Quality on School Outcomes and Improvement: A Value-Added Approach." *Educational Administration Quarterly* 36(4):513–552.

Heil, Scott, Liza Reisel, and Paul Attewell. 2014. "College Selectivity and Degree Completion." *American Educational Research Journal* pp. 913–935.

HEOA. 2008. "Higher Education Opportunity Act."

Hillman, Nicholas W., Alisa Hicklin Fryar, and Valerie Crespín-Trujillo. 2018. "Evaluating the Impact of Performance Funding in Ohio and Tennessee." *American Educational Research Journal* 55(1):144–170.

Hirano, Keisuke and Guido W Imbens. 2001. "Estimation of Causal Effects using Propensity Score Weighting: An Application to Data on Right Heart Catheterization." *Health Services & Outcomes Research Methodology* 2:259–78.

Hochschild, Jennifer L. and Nathan Scovronick. 2004. *The American Dream and the Public Schools*. New York, NY: Oxford University Press.

Hoekstra, Mark. 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.

Hout, Michael. 1988. "More Universalism, Less Structural Mobility: The American Occupational Structure in the 1980s." *American Journal of Sociology* 93(6):1358–1400.

Hout, Michael. 2012. "Social and Economic Returns to College Education in the United States." *Annual Review of Sociology* 38(1):379–400.

Huisman, Jeroen and Jan Currie. 2004. "Accountability in higher education: Bridge over troubled water?" *Higher Education* 48(4):529–551.

Indiana University Center for Postsecondary Research. n.d. "The Carnegie Classification of Institutions of Higher Education, 2015 Edition." Technical report, Bloomington, IN.

Jenkins, Davis and Fink, John. 2016. "Tracking Transfer: New Measures of Institutional and State Effectiveness in Helping Community College Students Attain Bachelor's Degrees." Technical report, Community College Research Center, Teachers College, Columbia University, New York, NY.

Jennings, Jennifer L., David Deming, Christopher Jencks, Maya Lopuch, and Beth E. Schueler. 2015. "Do Differences in School Quality Matter More Than We Thought? New Evidence on Educational Opportunity in the Twenty-first Century." *Sociology of Education* 88(1):56–82.

Jennings, Jennifer L. and Thomas A. DiPrete. 2010. "Teacher Effects on Social and Behavioral Skills in Early Elementary School." *Sociology of Education* 83(2):135–159.

Karabel, Jerome. 2006. *The Chosen: The Hidden History of Admission and Exclusion at Harvard, Yale, and Princeton*. Houghton Mifflin Harcourt.

Karen, David. 1990. "Toward a Political-Organizational Model of Gatekeeping: The Case of Elite Colleges." *Sociology of Education* 63(4):227–240.

Kelchen, Robert. 2018. *Higher Education Accountability*. Baltimore, MD: Johns Hopkins University Press.

Kelchen, Robert and Luke J. Stedrak. 2016. "Does Performance-Based Funding Affect Colleges' Financial Priorities?" *Journal of Education Finance* 41(3):302–321.

Khan, Shamus. 2012. *Privilege: The Making of an Adolescent Elite at St. Paul's School*. Princeton, N.J.: Princeton University Press.

Kirst, Michael and Mitchell L. Stevens. 2015. *Remaking College: The Changing Ecology of Higher Education*. Palo Alto, CA: Stanford University Press.

Kuh, George and Stanley Ikenberry. 2009. "More Than You Think, Less Than We Need: Learning Outcomes Assessment in American Higher Education." Technical report, University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment, Urbana, IL.

Kullback, Solomon. 1959. *Information Theory and Statistics*. New York, NY: John Wiley & Sons.

Kurlaender, Michal, Scott Carrell, and Jacob Jackson. 2016. "The Promises and Pitfalls of Measuring Community College Quality." *RSF* 2(1):174–190.

Kurlaender, Michal and Eric Grodsky. 2013. "Mismatch and the Paternalistic Justification for Selective College Admissions." *Sociology of Education* 86(4):294–310.

Labaree, David F. 1997. "Public Goods, Private Goods: The American Struggle Over Educational Goals." *American Educational Research Journal* 34(1):39–81.

Labaree, David F. 2012. *Someone Has to Fail*. Harvard University Press.

Lara-Millán, Armando. 2014. "Public Emergency Room Overcrowding in the Era of Mass Imprisonment." *American Sociological Review* 79(5):866–887.

Lee, Valerie E., Julia B. Smith, and Robert G. Croninger. 1997. "How High School Organization Influences the Equitable Distribution of Learning in Mathematics and Science." *Sociology of Education* 70(2):128–150.

Legewie, Joscha and Thomas A. DiPrete. 2012. "School Context and the Gender Gap in Educational Achievement." *American Sociological Review* 77(3):463–485.

Li, Amy Y. 2018. "Lessons Learned: A Case Study of Performance Funding in Higher Education." Technical report, Third Way.

Liu, Ou Lydia. 2011. "Value-added assessment in higher education: a comparison of two methods." *Higher Education* 61(4):445–461.

Lucas, Samuel R. 2001. "Effectively Maintained Inequality: Education Transitions, Track Mobility, and Social Background Effects." *American Journal of Sociology* 106(6):1642–1690. ArticleType: research-article / Full publication date: May 2001 / Copyright © 2001 The University of Chicago Press.

March, James G. and Johan P. Olsen. 1976. *Ambiguity and Choice in Organizations*. Bergen: Universitetsforlaget.

Massey, Douglas S., Camille Z. Charles, Garvey Lundy, and Mary J. Fischer. 2011. *The Source of the River: The Social Origins of Freshmen at America's Selective Colleges and Universities*. Princeton University Press.

Mehta, Jal. 2013. "How Paradigms Create Politics , How Paradigms Create Politics: The Transformation of American Educational Policy, 1980–2001 , The Transformation of American Educational Policy, 1980–2001." *American Educational Research Journal* 50(2):285–324.

Meyer, Heinz-Dieter and Brian Rowan (eds.). 2006. *The New Institutionalism in Education*. Albany, NY: State University of New York Press.

Meyer, John W. 1977. "The Effects of Education as an Institution." *American Journal of Sociology* 83(1):55–77.

Meyer, John W., Francisco O. Ramirez, David John Frank, and Evan Schofer. 2007. "Higher Education as an Institution." In *Sociology of Higher education: Contributions and Their Contexts*, edited by Patricia J. Gumport, pp. 187–221. Baltimore, MD: Johns Hopkins University Press.

Meyer, John W. and Brian Rowan. 1977. "Institutionalized Organizations: Formal Structure as Myth and Ceremony." *American Journal of Sociology* 83(2):340–363.

Meyer, John W. and Brian Rowan. 1978. "The Structure of Educational Organizations." In *Environments and Organizations*, edited by Marshall W. Meyer. San Francisco, CA: Jossey-Bass Inc.

Meyer, John W. and W. Richard Scott. 1983. "Centralization and the Legitimacy Problems of Local Government." In *Organizational Environments: Ritual and Rationality*, edited by John W. Meyer and W. Richard Scott, pp. 199–215. Beverly Hills: Sage Publications.

Meyer, Robert H. 1997. "Value-added indicators of school performance: A primer." *Economics of Education Review* 16(3):283–301.

Morgan, Stephen L. and Sol Bee Jung. 2016. "Still No Effect of Resources, Even in the New Gilded Age?" *RSF: The Russell Sage Foundation Journal of the Social Sciences* 2(5):83–116.

Morgan, Stephen L. and Christopher Winship. 2007. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge, UK: Cambridge University Press.

Morse, Robert and Matt Mason. 2017. "How U.S. News Collects Best Colleges Rankings." *US News* .

NCES. 2018. "Statutory Requirements for Reporting IPEDS Data."

Nichols, Andrew Howard, Kimberlee Eberle-Sudré, and Meredith Welch. 2016. "Rising Tide II: Do Black Students Benefit as Grad Rates Increase?" Technical report, The Education Trust, Washington, DC.

Oakes, Jeannie. 1985. *Keeping Track: How Schools Structure Inequality*. Birmingham, NY: Vail-Ballou Press.

Obama, Barack. 2009. "State of the Union Address." In *The American Presidency Project*, edited by John T. Wooley and Gerhard Peters. http://www.presidency.ucsb.edu/: University of California , Santa Barbara.

Peugh, James L. and Craig K. Enders. 2004. "Missing Data in Educational Research: A Review of Reporting Practices and Suggestions for Improvement." *Review of Educational Research* 74(4):525–556.

Posselt, Julie R. and Eric Grodsky. 2017. "Graduate Education and Social Stratification." *Annual Review of Sociology* 43(1):353–378.

Rabe-Hesketh, Sophia and Anders Skrondal. 2012. *Multilevel and Longitudinal Modeling Using Stata*. College Station, TX: Stata Press, 3rd edition.

Raudenbush, Stephen and Anthony S. Bryk. 1986. "A Hierarchical Model for Studying School Effects." *Sociology of Education* 59(1):1–17.

Raudenbush, Stephen W. and JDouglas Willms. 1995. "The Estimation of School Effects." *Journal of Educational and Behavioral Statistics* 20(4):307–335.

Reich, Adam Dalton. 2014. *Selling Our Souls: The Commodification of Hospital Care in the United States*. Princeton University Press.

Riegle-Crumb, Catherine, George Farkas, and Chandra Muller. 2006. "The Role of Gender and Friendship in Advanced Course Taking." *Sociology of Education* 79(3):206–228.

Roksa, Josipa, Eric Grodsky, Richard Arum, and Adam Gamoran. 2007. "United States: Changes in Higher Education and Social Stratification." In *Stratification in Higher Education: A Comparative Study*, edited by Yossi Shavit, Richard Arum, and Adam Gamoran, pp. 165–194. Stanford, California: Stanford University Press.

Rosenbaum, James E., Caitlin E. Ahearn, and Janet E. Rosenbaum. 2017. *Bridging the Gaps: College Pathways to Career Success*. New York, NY: Russell Sage Foundation.

Rosenbaum, James E., Regina Deil-Amen, and Ann E. Person. 2006. *After Admission: From College Access to College Success*. New York, NY: Russell Sage Foundation.

Rosenbaum, Paul R. and Donald B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70(1):41–55.

Rubin, Donald B. 1976. "Inference and Missing Data." *Biometrika* 63(3):581–592.

Rubin, Donald B. 2005. "Causal Inference Using Potential Outcomes." *Journal of the American Statistical Association* 100(469):322–331.

Scott, W. Richard. 2004. "Reflections on a Half-Century of Organizational Sociology." *Annual Review of Sociology* 30(1):1–21.

Scott, W. Richard. 2014. *Institutions and Organizations: Ideas and Interests*. SAGE, 4th edition.

Scott, W. Richard and John W. Meyer. 1988. "Environmental Linkages and Organizational Complexity: Public and Private Schools." In *Comparing Public and Private Schools*, edited by H.M Levin and T. James, pp. 128–60. New York, NY: Falmer Press.

Selznick, Philip. 1948. "Foundations of the Theory of Organization." *American Sociological Review* 13(1):25–35.

Shapiro, D., A. Dundar, X. Yuan, A. Harrell, J. Wild, and M. Ziskin. 2014. "Some College, No Degree: A National View of Students with Some College Enrollment, but No Completion." Signature report no. 7, National Student Clearinghouse Research Center.

Shin, Jung Cheol. 2010. "Impacts of performance-based accountability on institutional performance in the U.S." *Higher Education* 60(1):47–68.

Small, Mario L. and Christopher Winship. 2007. "Black students' graduation from elite colleges: Institutional characteristics and between-institution differences." *Social Science Research* 36(3):1257–1275.

Smith, Ashley A. 2017. "City Colleges of Chicago Faces Scrutiny Over Completion Rate Gains." *Inside Higher Education* November 8.

Snyder, Thomas D., Cristobal de Brey, and Sally A. Dillow. 2018. "Digest of Education Statistics 2016." Technical report, U.S. Department of Education, Washington, D.C.

Stevens, Mitchell L. 2009. *Creating a Class: College Admissions and the Education of Elites*. Cambridge, MA: Harvard University Press.

Stevens, Mitchell L., Elizabeth A. Armstrong, and Richard Arum. 2008. "Sieve, Incubator, Temple, Hub: Empirical and Theoretical Advances in the Sociology of Higher Education." *Annual Review of Sociology* 34(1):127–151.

Stevens, Mitchell L. and Ben Gebre-Medhin. 2016. "Association, Service, Market: Higher Education in American Political Development." *Annual Review of Sociology* 42(1):121–142.

Stuber, Jenny M. 2012. *Inside the College Gates: How Class and Culture Matter in Higher Education*. Lanham, MD: Lexington Books.

Tekwe, Carmen D., Randy L. Carter, Chang-Xing Ma, James Algina, Maurice E. Lucas, Jeffrey Roth, Mario Ariet, Thomas Fisher, and Michael B. Resnick. 2004. "An Empirical Comparison of Statistical Models for Value-Added Assessment of School Performance." *Journal of Educational and Behavioral Statistics* 29(1):11–36.

Thornton, Patricia H. and William Ocasio. 1999. "Institutional Logics and the Historical Contingency of Power in Organizations: Executive Succession in the Higher Education Publishing Industry, 1958 to 1990." *American Journal of Sociology* 105(3):801–843.

Torche, Florencia. 2011. "Is a College Degree Still the Great Equalizer? Intergenerational Mobility across Levels of Schooling in the United States." *American Journal of Sociology* 117(3):763–807.

Vaughan, Assistant Professor of Sociology Diane and Diane Vaughan. 1996. *The Challenger Launch Decision: Risky Technology, Culture, and Deviance at NASA*. University of Chicago Press.

Watkins-Hayes, Celeste. 2009. *The New Welfare Bureaucrats: Entanglements of Race, Class, and Policy Reform*. University of Chicago Press.

Weick, Karl E. 1976. "Educational Organizations as Loosely Coupled Systems." *Administrative Science Quarterly* 21(1):1–19.

Wilson, James Q. 1991. *Bureaucracy: What Government Agencies Do and Why They Do It*. New York, NY: Basic Books Classics.

Zemsky, Robert. 2010. "Accountability in the United States: Sorting through an American Muddle." In *Accountability in Higher Education: Global Perspectives on Trust and Power*, edited by Björn Stensaker and Lee Harvey, pp. 157–76. New York, NY: Routledge.

Zhao, Qingyuan and Daniel Percival. 2017. "Entropy balancing is doubly robust." *Journal of Causal Inference* 5(1).

Zhou, Xiang. 2019. "Equalization or Selection? Reassessing the "Meritocratic Power" of a College Degree in Intergenerational Income Mobility." *American Sociological Review* .

A Appendix Tables

Table A.1: Descriptive Statistics

	All Students, Including Two-to-Four-year Transfers (*n=126,455*)		Four-year Beginners (*n=81,985)*	
	Mean	Std. Dev.	Mean	Std. Dev.
Individual-level Variables				
BA completion within initial college	0.513		0.487	
Entry age	19.554	3.971	18.972	2.983
Female	0.591		0.578	
White	0.305		0.325	
Black	0.249		0.213	
Hispanic	0.255		0.250	
Asian	0.191		0.213	
U.S. citizen	0.692		0.731	
Permanent resident	0.205		0.187	
Visa-holder / refugee	0.044		0.036	
Undocumented	0.039		0.039	
Unknown	0.019		0.007	
Pell Grant recipient	0.521		0.494	
State funding recipient	0.591		0.606	
Family income	44,972.27	46,373.99	48,937.41	51,115.70
Economic disadvantage flag	0.783		0.764	
Eligible for low-income, high-talent program	0.109		0.144	
Dependent	0.906		0.948	
Total financial award offered upon application	4,189.409	3,085.406	4,483.18	3,287.36
From same city as college	0.816		0.806	
From same state as college	0.065		0.082	
From outside the college state	0.040		0.042	
From abroad	0.079		0.070	
Live within same geographical boundaries as college	0.352		0.361	
Speak limited English	0.285		0.290	
Disabled	0.020		0.015	
Single parent	0.030		0.017	
Attended public high school in same city as college	0.638		0.652	

Table A.1: Descriptive Statistics

Enrolled within 15 months of high school graduation	0.819		0.873	
High school GPA	—		82.487	6.577
Number of college credits earned in high school	3.794	0.889	4.033	0.740
State English test score in 12th grade	74.743	13.259	79.290	10.295
State math test score in 12th grade	76.378	13.931	80.291	11.697
Number of MetroU colleges to which applied	3.347	2.326	3.793	2.138
Enrolled in four-year college in September	0.859		0.901	
Enrolled in four-year college in January	0.141		0.099	
First enrolled in MetroU system in 1999	0.030		—	
First enrolled in MetroU system in 2000	0.047		—	
First enrolled in MetroU system in 2001	0.121		0.112	
First enrolled in MetroU system in 2002	0.135		0.129	
First enrolled in MetroU system in 2003	0.137		0.138	
First enrolled in MetroU system in 2004	0.141		0.147	
First enrolled in MetroU system in 2005	0.134		0.146	
First enrolled in MetroU system in 2006	0.128		0.153	
First enrolled in MetroU system in 2007	0.119		0.163	
First enrolled in MetroU system in 2008	0.008		0.013	
Transfer student	0.352		0	
College-Level Variables				
Transfer-out rate for four-year beginners	—		0.338	
Total average transfer-out rate	0.291		—	
Total average students per college	13,085.06	3,698.68	13,085.06	3,698.68
Proportion underrepresented minority	0.504		0.495	
Proportion receiving Pell Grants	0.521		0.519	

Table A.1: Descriptive Statistics

Proportion attending college with "mid-selective" Barron's rating	0.712		0.784	
Proportion "Less Selective" in Barron's rating	0.288		0.216	

Notes: Proportions may not add to one due to rounding. High school GPA omitted for the full sample due to differences in reporting for for freshmen and transfers.
Source: MetroU administrative data.

Table A.2: Logistic Regression Results using the WMA and ENT Approaches, as Log Odds

	WMA		ENT	
	Coef.	*Std. Err.*	*Coef.*	*Std. Err.*
College 1	0.199***	0.033	0.224***	0.039
Entry age	0.008*	0.004	-0.017	0.011
Female	0.282***	0.016	0.226***	0.024
U.S. citizen	0.084*	0.014	0.061	0.022
Black	-0.088**	0.032	-0.053	0.052
Hispanic	-0.279***	0.032	-0.247***	0.046
Asian	-0.048	0.032	0.006	0.042
Pell Grant recipient	-0.325***	0.037	-0.195***	0.053
Black*Pell	-0.035	0.045	-0.125	0.078
Hispanic*Pell	-0.012	0.044	-0.132	0.068
Asian*Pell	0.012	0.046	-0.006	0.060
State funding recipient	0.025	0.025	0.047	0.039
Family income	0.178***	0.011	0.310***	0.021
Economic disadvantage flag	0.495***	0.028	0.402***	0.042
Eligible for low-income, high-talent program	-0.045	0.027	0.010	0.465
Dependent	0.210***	0.042	0.124	0.080
Total financial award offered upon application	0.373***	0.014	0.310***	0.021
Live in same state as college	-0.032	0.032	0.007	0.047
Live outside of the college state, in the U.S.	-0.110*	0.043	-0.098	0.058
Live outside of the U.S.	0.262***	0.045	0.210***	0.067
Live within 100 miles of college	0.116***	0.017	0.081**	0.029
Speak limited English	0.030	0.019	0.054*	0.028
Disabled	0.308***	0.063	0.226*	0.099
Attended public high school	0.112***	0.021	0.201***	0.033
Enrolled within 15 months of high school graduation	0.275***	0.029	0.249***	0.052
High school GPA	0.429***	0.011	0.451***	0.017
High school GPA squared	0.041***	0.007	0.049***	0.012
Number of college credits earned in high school	0.157***	0.010	0.161***	0.015
State English test score in 12th grade	0.046***	0.010	0.039*	0.015
State math test score in 12th grade	-0.024*	0.010	0.043**	0.016
Applied to one MetroU college	0.086*	0.036	-0.010	0.059
Applied to two MetroU colleges	0.258*	0.125	-0.105	0.064
Applied to three MetroU colleges	0.725**	0.210	-0.161**	0.059
Applied to four MetroU colleges	0.907***	0.207	-0.172**	0.059
Applied to five MetroU colleges	0.882***	0.207	-0.229***	0.063
Applied to six MetroU colleges	0.838***	0.206	-0.298***	0.053
Enrolled in September	0.154***	0.031	0.241***	0.053
Enrolled in 2002	-0.043	0.031	-0.019	0.046
Enrolled in 2003	-0.113***	0.032	-0.064	0.046
Enrolled in 2004	-0.105**	0.032	-0.038	0.047
Enrolled in 2005	-0.080*	0.032	0.028	0.047
Enrolled in 2006	-0.082**	0.031	0.011	0.047

Table A.2: Logistic Regression Results using the WMA and ENT Approaches, as Log Odds

Enrolled in 2007	-0.101**	0.031	0.001	0.047
Enrolled in 2008	-0.111	0.079	0.174	0.142
Transfer-out rate	-3.049***	0.031	-2.662***	0.294
WMA Groupings (4,149 total groups)	X	X	—	—
Constant	-0.253	0.137	0.113	0.305

Notes: Omitted category is "white" for race, "live in same city" for place of residence, "0" for college application, "January" for enrollment month, and 2001 for enrollment year. *p<0.05; **p<0.01; ***p<0.001.
Source: MetroU Administrative Data

Table A.3: Alternative Within-Matched-Applicant Measures of College Effects on Ten-year, Within-College BA Completion Rates for Four-year Beginners

	Logistic Regression Results				Predicted Marginal Effects Differences			
	Only Matched Applicants		Alternative Matching		Only Matched Applicants		Alternative Matching	
	Odds Ratio	Standard Error	Odds Ratio	Standard Error	Distance from Counter-factual Mean (percent-age points)	Chi-square test value	Distance from Counter-factual Mean (percentage points)	Chi-square test value
College 1	1.383***	0.048	1.265***	0.041	6.4	0.000	4.9	0.000
College 2	0.904**	0.043	0.850***	0.035	-2.1	0.017	-3.4	0.000
College 3	0.864*	0.067	0.854***	0.056	-3.1	0.029	-3.5	0.004
College 4	0.884**	0.025	0.945	0.033	-2.5	0.004	-1.2	0.082
College 5	1.075	0.041	1.012	0.031	1.5	0.077	0.2	0.714
College 6	1.477***	0.057	1.291***	0.040	8.4	0.000	5.5	0.000
College 7	0.966	0.042	1.068*	0.032	-0.7	0.399	1.4	0.043
College 8	0.873*	0.060	0.981	0.041	-2.9	0.022	-0.4	0.647
College 9	0.945	0.063	1.044	0.047	1.0	0.373	0.8	0.367
College 10	0.672**	0.141	0.676***	0.078	-6.7	0.003	-6.8	0.000
College 11	0.475***	0.122	0.567***	0.098	-12.7	0.000	-10.0	0.000

Notes: For the logistic regression coefficients, *p<0.05; **p<0.01; ***p<0.001.
Control variables include student race, family income, Pell Grant status, the interaction between race and Pell Grant status, state grant aid status, special college program status, age, sex, citizenship status, dependent status, single parent status, place of permanent resident, total financial aid awarded prior to four-year college entry, whether or not a student is disabled, whether or not a student speaks proficient English, whether students attended a public or a private high school, whether the college geographically proximate to a student's high school, whether or not a student started college in the 15 months following high school graduation, high school GPA, squared high school GPA, math state test scores, English state test scores, number of college-applicable high school credits, number of MetroU colleges to which the studies applied, the year and month of college entry, and average college transfer out rates.
Source: MetroU administrative data.

Table A.4: Pairwise, Percentage Point Differences in College Effectiveness using Ten-year, Within-College BA Completion Rates for Four-year Beginners

	College 1	College 2	College 3	College 4	College 5	College 6	College 7	College 8	College 9	College 10	College 11
College 1	—	5.4 (0.000)	9.1 (0.001)	9.8 (0.000)	8.1 (0.000)	3.4 (0.002)	7.1 (0.000)	9.8 (0.000)	12.5 (0.000)	23.0 (0.000)	17.1 (0.000)
College 2	—	—	6.5 (0.000)	2.9 (0.000)	4.5 (0.000)	2.0 (0.173)	3.4 (0.001)	2.5 (0.096)	10.5 (0.000)	12.4 (0.000)	13.3 (0.000)
College 3	—	—	—	0.8 (0.525)	-3.3 (0.174)	-2.3 (0.051)	-4.5 (0.092)	2.2 (0.209)	8.4 (0.000)	8.5 (0.001)	14.1 (0.007)
College 4	—	—	—	—	0.3 (0.658)	-0.6 (0.637)	-0.3 (0.694)	-2.2 (0.076)	7.3 (0.000)	12.3 (0.000)	11.7 (0.000)
College 5	—	—	—	—	—	-5.0 (0.000)	-0.2 (0.798)	0.3 (0.837)	3.9 (0.050)	11.7 (0.000)	8.2 (0.000)
College 6	—	—	—	—	—	—	4.2 (0.000)	3.7 (0.000)	13.8 (0.000)	6.5 (0.000)	15.1 (0.000)
College 7	—	—	—	—	—	—	—	0.6 (0.556)	6.8 (0.000)	7.7 (0.000)	12.4 (0.000)
College 8	—	—	—	—	—	—	—	—	6.0 (0.000)	7.0 (0.002)	11.4 (0.000)
College 9	—	—	—	—	—	—	—	—	—	2.2 (0.111)	8.2 (0.000)
College 10	—	—	—	—	—	—	—	—	—	—	6.5 (0.009)

Notes: College effects are estimated using the ENT approach. Values are expressed as percentage point deviations from the counterfactual mean, or the expected BA completion rate while attending the other college in the pairwise comparison. Chi-square test values are included in parentheses. Control variables include student race, family income, Pell Grant status, the interaction between race and Pell Grant status, state grant aid status, special college program status, age, sex, citizenship status, dependent status, single parent status, place of permanent resident, total financial aid awarded prior to four-year college entry, whether or not a student is disabled, whether or not a student speaks proficient English, whether students attended a public or a private high school, whether the college geographically proximate to a student's high school, whether or not a student started college in the 15 months following high school graduation, high school GPA, squared high school GPA, math state test scores, English state test scores, number of college-applicable high school credits, number of MetroU colleges to which the studies applied, the year and month of college entry, and average college transfer out rates.
Source: MetroU administrative data.

Table A.5: Heterogeneity of Between-College Effects by Student Race and Income among Four-year Beginners, Expressed as Odds Ratios

	College 1	College 2	College 3	College 4	College 5	College 6	College 7	College 8	College 9	College 10	College 11
White	0.944	1.115*	0.969	1.044	0.950	1.308***	0.954	0.927	0.752	0.599**	4.035
	(0.061)	(0.053)	(0.064)	(0.042)	(0.044)	(0.070)	(0.060)	(0.090)	(0.130)	(0.099)	(4.724)
Black	1.133	0.806*	0.480*	1.114*	0.869*	1.630***	0.996	0.911	1.265**	0.664***	0.521***
	(0.114)	(0.072)	(0.140)	(0.059)	(0.048)	(0.093)	(0.047)	(0.049)	(0.105)	(0.071)	(0.059)
Hispanic	1.099	0.945	0.897	0.850*	1.007	1.450***	0.945	1.048	0.656***	0.740**	0.141***
	0.085	(0.063)	(0.134)	(0.056)	(0.049)	(0.066)	(0.040)	(0.044)	(0.065)	(0.081)	(0.542)
Asian	1.953***	0.701***	0.516***	0.825***	1.198**	1.392***	0.985	0.755**	0.913	0.893	0.074*
	(0.143)	(0.039)	(0.073)	(0.045)	(0.064)	(0.125)	(0.053)	(0.081)	(0.124)	(0.134)	(1.051)
Pell Recipient	1.428***	0.807***	0.614***	0.946	0.961	1.403***	0.940	1.035	1.125	0.728***	0.561***
	(0.087)	(0.036)	(0.060)	(0.033)	0.034	(0.052)	(0.031)	(0.038)	(0.079)	(0.060)	(0.076)
Non-Pell Recipient	1.150**	0.973	0.966	0.994	1.023	1.419***	0.990	0.835***	1.088	0.745**	0.353***
	(0.059)	(0.039)	(0.062)	(0.035)	(0.035)	(0.059)	(0.037)	(0.042)	(0.095)	(0.072)	(0.059)

Notes: College effects are estimated using the ENT approach, where the "treatment" is attendance of one of the eleven MetroU colleges. Standard errors are captured in parentheses, *p<0.05; **p<0.01; ***p<0.001.
Control variables include family income, student race (in the Pell models), Pell Grant status (in the race models), state grant aid status, special college program status, age, sex, citizenship status, dependent status, single parent status, place of permanent resident, total financial aid awarded prior to four-year college entry, whether or not a student is disabled, whether or not a student speaks proficient English, whether students attended a public or a private high school, whether the college geographically proximate to a student's high school, whether or not a student started college in the 15 months following high school graduation, high school GPA, math state test scores, English state test scores, number of college-applicable high school credits, number of MetroU colleges to which the studies applied, and the year and month of college entry.
Source: MetroU administrative data.

Table A.6: Heterogeneity of Within-College Effects by Student Race and Income among Four-year Beginners, Expressed as Odds Ratios

	College 1	College 2	College 3	College 4	College 5	College 6	College 7	College 8	College 9	College 10	College 11
White	—	—	—	—	—	—	—	—	—	—	1.326
											(1.443)
Black	0.927	0.658***	0.481**	0.950	0.846*	1.123	0.952	0.735	1.343	1.065	—
	(0.085)	(0.056)	(0.116)	(0.066)	(0.058)	(0.103)	(0.096)	(0.125)	(0.251)	(0.266)	
Hispanic	0.648***	0.614***	0.751	0.585***	0.831**	0.865	0.584***	0.786	0.725	0.660	0.210**
	(0.053)	(0.043)	(0.111)	(0.055)	(0.057)	(0.080)	(0.064)	(0.169)	(0.131)	(0.193)	(0.114)
Asian	1.080	0.710***	0.657*	0.803*	1.092	1.102	0.912	0.798	1.039	0.712	0.015***
	(0.068)	(0.064)	(0.026)	(0.076)	(0.063)	(0.123)	(0.090)	(0.178)	(0.195)	(0.192)	(0.012)
Pell Recipient	1.012	0.501**	1.072	1.236	1.152	0.824	0.912	2.090	0.909	1.004	1.489
	(0.221)	(0.117)	(0.304)	(0.279)	(0.202)	(0.217)	(0.201)	(0.574)	(0.226)	(0.306)	(0.673)

Notes: College effects are estimated using the ENT approach. Values are expressed as odds ratios where white students are the omitted group in racial comparisons and the non-Pell mean in income comparisons. College 11 is an exception in that contrasts are made with black students. Standard errors are captured in parentheses, *p<0.05; **p<0.01; ***p<0.001.
Control variables include family income, student race (in the Pell models), Pell Grant status (in the race models), state grant aid status, special college program status, age, sex, citizenship status, dependent status, single parent status, place of permanent resident, total financial aid awarded prior to four-year college entry, whether or not a student is disabled, whether or not a student speaks proficient English, whether students attended a public or a private high school, whether the college geographically proximate to a student's high school, whether or not a student started college in the 15 months following high school graduation, high school GPA, math state test scores, English state test scores, number of college-applicable high school credits, number of MetroU colleges to which the studies applied, and the year and month of college entry.
Source: MetroU administrative data.

Table A.7: The Effects of College-level Characteristics on BA Completion for Four-year Beginners, as Odds Ratios

	Coef.	*Std. Err.*
Transfer-out Rate (std)	0.893	0.064
Size (std)	1.340*	0.130
Percent Underrepresented Minority (std)	0.903	0.157
Percent Pell Grant recipient (std)	1.121	0.115
Mid-Selective rating	0.817	0.218
Random Intercept (std. dev.)	0.137***	0.031
Individual-level controls	X	
Month, year, and application controls	X	
Constant	0.327***	0.198

Notes: The omitted category for "mid-selective rating" is non-selective rating.
*p<0.05, **p<0.01, ***p<0.001.
Source: MetroU administrative data.