“Workhorses of Opportunity”: Regional Universities Increase Local Social Mobility*

Greg Howard† Russell Weinstein‡

March 5, 2025

Abstract

Regional public universities educate approximately 70 percent of college students at four-year public universities and an even larger share of students from disadvantaged backgrounds. They aim to provide opportunity for education and social mobility, in part by locating near potential students. In this paper, we use the historical assignment of normal schools and insane asylums (normal schools grew into regional universities while asylums remain small) and data from Opportunity Insights to identify the effects of regional universities on the social mobility of nearby children. Children in counties given a normal school get more education and have better economic and social outcomes, especially lower-income children. For several key outcomes, we show this effect is a causal effect on children, and not only selection on which children live near universities. We use student-level survey data to compare characteristics of college-going students from normal school and asylum counties and to study the geographic barriers that keep asylum-county children from attending college.

JEL Codes: J62, I23, I26, R53

*We would like to thank Andy Garin, as well as the seminar and conference participants at the University of Illinois; the University of Delaware; Iowa State; Grinnell College; the University of California, Berkeley; the Queen's-Toronto Labor Workshop on the Economics of Inequality and Discrimination; the Association for Education Finance and Policy annual conference; the Federal Reserve Bank of Chicago; the Mid-Midwest Applied Microeconomics Conference at Purdue; the University of Warwick; Miami University; the Upjohn Institute; the IZA/ECONtribute Workshop on the Economics of Education; the LERA@ASSA 2024 Conference; Einaudi Institute for Economics and Finance; Luiss University; the European University Institute, and Michigan State University for helpful comments. Greg would like to thank the Upjohn Institute for helping to fund the project with an Early Career Research Award.

†University of Illinois, Urbana-Champaign. glhoward@illinois.edu

‡University of Illinois, Urbana-Champaign. weinst@illinois.edu

Regional public universities have been considered the "colleges of the forgotten Americans" and "workhorses of opportunity" because of their potential to increase social mobility (Dunham, 1969; Wendler, 2018). From their establishment in the mid-20th century, a central part of their mission has been to increase access to higher education, by locating near potential students, being less selective, and having lower tuition. Regional public universities enroll roughly 40 percent of all undergraduate students in the United States, and students at these institutions come disproportionately from lower-parental-income families and are more likely to be racial minorities relative to other four-year public universities.[1]

We study whether regional public universities increase nearby children's educational attainment and social mobility, using data from Opportunity Insights. These local objectives are an important justification for these regional colleges, making our analysis especially policy-relevant. Our findings highlight the continued role of geographic frictions in college attendance and economic mobility.

The central identification challenge is that universities are not located randomly, and they may have been placed in areas expected to have high educational attainment and economic mobility even without the university. We use a strategy developed in Howard, Weinstein and Yang (2024) to identify the impact of regional public universities on the county. Our strategy utilizes the placement of normal schools and insane asylums in the late 19th and early 20th centuries, both part of the period's social reform movements. We show that state governments assigned these institutions to counties using similar criteria, including

[1]Of undergraduates at four-year public universities, almost 70 percent, including 85 percent of Black and 74 percent of Hispanic students, are at regional universities (Fryar, 2015). These statistics are based on Fryar (2015)'s historical definition of comprehensive universities, which includes public, four-year universities that are not the primary research university in the state, not land-grant universities, and not established expressly to serve as a research institution.

On average roughly 15 percent of students at four-year public flagship, public elite, or public highly selective institutions came from the bottom two parental income quintiles. At four-year public non-flagship, non-highly selective institutions that fraction is roughly 28 percent. This is based on the university-level Opportunity Insights, and the 1980 birth cohort (Chetty et al., 2020). We exclude institutions that report as a system. There are 45 public institutions in the flagship/highly selective group, and 377 in the other (314 of which were classified as selective public colleges and the remainder were nonselective public colleges). We note that the distinctions used for this statistic are not quite the same as Fryar (2015)'s historical definition, but nonetheless separates the very selective and flagships from other public universities. Pell grants are also more common at regional universities (Maxim and Muro, 2020).

political factors, proximity and ease of access to population centers, as well as locations with sufficient property and natural beauty (Humphreys, 1923; Kirkbride, 1854). By the mid-20th century, normal schools had evolved to become regional public universities, and they comprise roughly half of today's regional public universities. In contrast to the universities, most insane asylums were converted into psychiatric health facilities and remain small in size. Our central identification assumption is that the asylum counties are a good counterfactual for what would have happened in the normal school counties had they instead received a different state institution that remained similar in size.

Under our identification assumption, there are two ways in which our specification could have a zero coefficient: either if universities have no effect on the social mobility of nearby children; or if universities impact local children, but outcomes are similar in asylum counties because students can easily travel within a state for college or because they can access local private institutions with similar effects. Thus, a non-zero coefficient rejects both the idea that universities have no impact and that geographic sorting or other universities render regional university locations irrelevant.

First, using the Census Tree (Price et al., 2023*b*), we show that children growing up in normal school counties in 1920, and reaching college-age in the 1920s and early 1930s, were more likely to have a college degree by 1940 relative to children growing up in asylum counties. This was a period when normal schools had become teachers colleges that awarded bachelor's degrees in education, but before the colleges had increased dramatically in size.

Next, we show that normal school assignment increases economic and social mobility in the county for the 1978-1983 birth cohorts from Chetty et al. (2018), a time when the normal schools had become regional public universities. As they did for children in 1920, these universities increase the fraction of children in the county who obtain at least a four-year degree and at least some college. The largest percentage increases are for children of lower-income parents. They also increase the high school graduation rate.

In addition, these universities improve the fraction of children in the county who are employed in their mid-30s as well as their income percentiles, with effects concentrated among children from lower-income families. Finally, we also see that regional public universities have impacts on social outcomes of children in their county: increasing the fraction of lower-income children in the county who get married and decreasing the fraction that live in their childhood commuting zone.

Using estimates from Chetty and Hendren (2018), we see evidence that these causal impacts on the county reflect causal impacts on individuals, rather than reflecting sorting of high mobility individuals into counties with regional public universities.

There are several channels through which normal school assignment could affect local economic mobility today. Many of our findings are consistent with decreased geographic frictions in college attendance. A back-of-the-envelope calculation suggests that our estimated effects on wages and on college degree attainment are consistent with the estimated returns to graduating from college found in the literature. In contrast, we do not find results consistent with the normal schools having large effects on the local economy, which then affects social mobility. In particular, county-level characteristics, such as the sectoral composition and income distribution, are similar between normal school and asylum counties, even today.[2]

Finally, we supplement our analysis using rich student-level data from The Freshman Survey, which includes 2.5 million individuals who grew up in normal school or asylum counties. We show evidence consistent with regional public universities inducing enrollment of local students that are on the margin of going to college. Furthermore, we use The Freshman Survey to explore the self-reported geographic frictions that may be preventing regional universities from reaching students in asylum counties. We see evidence consistent with proximity to a regional university reducing financial costs of college, as well as reducing information frictions.

[2]A reader may wonder how a regional university would improve social outcomes for children without improving the overall local economy. We find that children from normal school counties are also more likely to move away from home as adults, so the benefits of the increased social mobility do not necessarily lead to higher local incomes or other social outcomes.

Despite their size and potential to improve economic mobility, the impacts of regional public universities have received limited attention in the literature, especially relative to community colleges and elite universities (Schneider and Deane, 2015).[3] While their role as an anchor institution in local communities is often cited, along with their role in enhancing mobility, there is little work to our knowledge estimating the causal impacts of these public, less research-intensive universities on nearby residents. This analysis is particularly relevant given recent discussions about consolidation and the future of these universities, such as in Pennsylvania, Vermont, and Wisconsin (see McClure and Fryar, 2020; Maxim and Muro, 2020; Seltzer, 2019, 2020; Quinton, 2020).

The relationship between educational attainment and proximity to universities has been an important topic in the literature.[4] For example, Card (1993) finds that proximity to a college raises education and earnings for men in the 1960s and 1970s, especially for men with the lowest predicted levels of educational attainment. [5] Our paper contributes to this literature in several ways. First, we use a novel strategy to identify the causal impact of universities on local educational attainment. Second, we focus on regional public universities, an understudied higher education sector. Finally, we utilize the rich data from the U.S. Census and the IRS made available by Opportunity Insights, allowing us to study the impacts on education for roughly 20 percent of the U.S. population born between 1978 and 1983, and other labor market and social outcomes for nearly the whole population. Many of the previous papers have used samples from survey data, such as the National Longitudinal

[3]Klor de Alva (2019) documents rates of income mobility among enrollees of comprehensive universities; Crisp, McClure and Orphan (2021) present a volume exploring broadly accessible institutions, including but not limited to regional public universities.

[4]More broadly our paper contributes to research studying universities and local economic growth, with many of the papers focusing on innovation. Papers include Aghion et al. (2009); Andersson, Quigley and Wilhelmson (2004); Andrews (2021); Bartik and Erickcek (2008); Cantoni and Yuchtman (2014); Feng and Valero (2020); Hausmann (2020); Kantor and Whalley (2014, 2019); Moretti (2004); Valero and Reenen (2019). Also related to our paper, Garin and Rothbaum (2022) study the long-run effects on economic mobility from counties receiving a large manufacturing plant in World War II.

[5]Other papers studying the relationship between proximity to universities and enrollment or completed education include: Bedard (2001); Kling (2001); Do (2004); Doyle and Skinner (2016); Kane and Rouse (1995); Long (2004); Jepsen and Montgomery (2009); Alm and Winters (2009); Mountjoy (2022). Mountjoy (2022) focuses on community colleges.

Surveys or the High School and Beyond survey.

Russell, Yu and Andrews (2022) and Russell and Andrews (2022), building on the empirical strategy of Andrews (2021), also focus on identifying the causal impact of colleges on educational attainment and economic mobility, by comparing areas with universities to runners-up locations for universities. Compared to our findings, they find larger effects on college education and smaller effects on income rank. We view our papers as complementary. One of the biggest differences is that we identify the effects of regional public universities, while the sample in Russell, Yu and Andrews (2022) includes primarily research-intensive universities. Given that regional public universities were established to improve local access to higher education and opportunity, ours is an especially relevant sample for understanding the impact of universities on mobility. Second, given the relatively few number of observations inherent to either empirical strategy, bringing more observations to this question is of particularly high return.[6] Finally, the counties in our control group are given a similarly-sized state institution, rather than being only runners-up. Russell, Yu and Andrews (2022) are also interested in the effect of universities relative to counties with a "consolation prize," but have only 27 counties in the sample for this exercise.

Finally, Chetty et al. (2014) and Chetty and Hendren (2018) show some evidence of a positive relationship between their local mobility measures and colleges per capita or the graduation rate at local colleges, though as Chetty and Hendren (2018) caution, this does not identify the effect of colleges on local mobility.

1 History of Normal Schools and Asylums

The social reform movements of the 19th century included support for public institutions aimed at societal improvement. These institutions included normal schools to train teachers and asylums to treat those with mental illnesses (Grob, 2008). In this section, we provide

[6] There are 191 counties in Russell, Yu and Andrews (2022), split between 63 counties with universities and 128 without. Our paper has 204 counties that received normal schools and 125 that received asylums.

qualitative evidence that locations for these institutions were chosen based on similar criteria. For a more detailed discussion, please see Howard, Weinstein and Yang (2024).

The original purpose of normal schools was to train teachers to meet growing demand stemming from the common school movement in the mid 19th century (Labaree, 2008). There were 209 state normal schools opened between 1839 and 1930 (Ogren, 2005). Similarly, as part of the mid-19th century movement to improve care for those with mental illnesses, many states opened insane asylums. The objective of these asylums was to facilitate recovery and to provide compassionate care (Grob, 2008).

The criteria for where to locate normal schools and insane asylums were very similar. Both were political decisions, in which population, geographic accessibility, and natural beauty were important factors. Humphreys (1923) describes in detail the location decisions for normal schools, asserting that political factors were the most important, though other factors included demand for instruction (e.g. local population), geographic accessibility, financial and land donations, location of existing schools, and natural beauty. Kirkbride (1854) developed an architectural plan for asylums, implemented by many states, which emphasized the importance of accessibility to population centers, locations with natural beauty, ample area for recreation, and stately architecture.[7]

During this period local communities desired and took pride in both types of institutions. An article from the *Kankakee Gazette*, written in August 1877 when the city was assigned an asylum, helps illustrate these points, “Our citizens received the news in a spirit of jubilee, and on Friday evening there was a bonfire, band music... and speeches...” The article expresses gratitude for “the great services of Messrs. Bonfeid and Taylor, our representatives in the upper and lower houses of the legislature,” highlighting the importance of the political process in determining these locations.

[7]Appendix A.2 shows states with higher income per capita in 1929 had built more normal schools and asylums ($p < .05$). Magnitudes suggest a positive relationship between state income per capita in 1929 and the number of normal schools built, and that this relationship does not weaken when looking at the number of public non-research universities in 1987, when there had been convergence of per capita income across states. However, these results are not precisely estimated. See Appendix A.2 for details.

As we show in Howard, Weinstein and Yang (2024), states determined locations for normal schools and insane asylums at roughly the same time (see Figure A1a). The timing and the similar selection criteria, along with individual state histories, support the idea that whether a community received a normal school or an asylum may have been effectively random. We showed in Howard, Weinstein and Yang (2024) that in the early 20th century, enrollment at the previous normal schools and the population in asylums were similar relative to county population (see Figure A1b). This provides further supportive evidence that being selected as the location for these two types of institutions may have required similar political influence, as the institutions may have been expected to confer similar advantages.

We support our identification assumption with several additional observations. First, roughly 17 percent of counties that were assigned asylums also were assigned normal schools (13 percent of normal school counties had asylums). This suggests similar selection criteria for the two types of institutions.[8] Finally, we note that asylum counties were often runners-up locations for public colleges and universities, as documented in Andrews (2021).[9] And in the opposite direction, normal school counties were often considered contenders for asylums, such as in Bloomington, IL.

1.1 Subsequent Evolution

Demand for higher education increased over the course of the 20th century, and normal schools evolved with these changes. In the early 20th century, many were renamed as teachers colleges, allowing them to confer bachelor's degrees in education. In the mid-20th century there was growing demand for degrees that did not focus on teacher training. Many policy discussions at the time focused on improving access through geographic accessibility (Doyle

[8]How much overlap we would expect depends on the distribution of political power and constraints on dividing the institutions across counties, in addition to the similarity of criteria. However, the fact that we do see overlap is suggestive that the criteria were similar.

[9]Of the 62 high-quality public college site selection experiments in Andrews (2021), 18 had runners-up that were asylum counties, although most of these experiments were for land grant institutions. Andrews (2021) discusses consolation prizes, and argues that assignment of one type of institution versus another was "as good as random".

and Skinner, 2016; Mayhew, 1969; Willingham, 1970; Douglass, 2007). Mayhew (1969) presents a summary of state master plans for higher education developed during this period of increased demand, stating "all plans seek to provide complete geographical access to higher education."

Many proponents thought normal schools should offer bachelor's degrees in areas other than education. Proponents argued they were uniquely positioned to increase access to a college education for their local areas because they already existed as higher education institutions and they were geographically distributed around states. For example, in advocating they be permitted to grant liberal arts degrees, college leaders at Eastern Illinois State Teachers College cited the limited number of other colleges in the region, that they were already serving as a regional college, and that many highly qualified high school students were not willing to attend a teachers college but would attend a state college (Coleman, 1950).[10]

The proponents of these changes were successful, and in the mid-20th century many of the teachers colleges were given the authority to grant degrees in areas other than education. As a result, many of the teachers colleges were renamed as state colleges.[11] In contrast to state universities, these state colleges focused on undergraduate education, business, teaching, and engineering (as opposed to law, medicine, and scholarship) (Mayhew, 1969). From the 1950s through the 1970s, many obtained university status (Labaree, 2008). Commenting on the frequent name changes, Dunham (1969) humorously noted that college stores would discount t-shirts with the college's previous name. Figure A1a documents the timing of these changes and Figure A1b shows large enrollment increases as normal schools converted to regional

[10]Similarly, proponents of making these changes at Southern Illinois Normal University argued local high school students were demanding a liberal arts degree, and it would be very costly for them to obtain this degree from another college (Lentz, 1955). A 1945 commission report wrote that even though they were only authorized to prepare teachers, the teachers colleges in Illinois had effectively already become regional colleges. These colleges were under pressure from the region to provide broader training, and students were enrolling in the teachers colleges and then not entering the teaching profession. The report noted that over the past seven years approximately 25% of graduates did not enter teaching (Commission to Survey Higher Educational Facilities in Illinois, 1945).

[11]Dunham (1969) observed that while many teachers colleges were renamed state colleges, they still remained focused on teacher training as of 1969. He also noted that for some faculty, "*teachers college* carries with it connotations of mediocrity, especially since Sputnik", and this led some faculty to push for removing "teachers" from the name of their college.

universities.

Institutions that started as normal schools comprise a large fraction of today's regional public universities, or using a similar classification, "comprehensive" universities (See Maxim and Muro (2020) for an overview of various classifications). Of the 320 public colleges in 1987 that are classified as "comprehensive" based on the 1987 Carnegie classification, roughly 50 percent started as state normal schools.[12] In keeping with their original mission, students at regional public universities are more likely to be from historically underrepresented or nontraditional groups in higher education (Fryar, 2015).

While some of the asylum buildings are no longer in use, states continue to own many of the asylum properties, and many are still used as psychiatric health facilities. Some properties are used as correctional facilities, while other have been acquired by universities (Hoopes, 2015). During the deinstitutionalization movement in the mid-20th century, institutionalized population per capita in asylum counties fell, though only modestly, and was twice the level in normal school counties in 2010 (see Figure A1b).

1.2 Data on Normal Schools and Asylums

As we describe in Howard, Weinstein and Yang (2024), we obtain data on normal schools' locations, opening years, and years corresponding to name changes, from Ogren (2005). Using the city and state of the normal school, we identified the county using StatsAmerica (Indiana Business Research Center, 2020). There were 209 normal schools across 204 counties, opened between 1839 and 1930, with median opening year of 1891 (Figure A1a). We digitize data on asylums' geographic locations and opening years from the 1923 special

[12] This is based on the evolution of name changes of state normal schools in Ogren (2005). In 1987, there are a total of 188 colleges that originated as state normal schools, based on Ogren (2005). Of these, 156 are classified as "comprehensive" in the 1987 Carnegie classifications, and 187 are "Research II," "Doctorate-Granting," "Comprehensive," or "Liberal Arts." Using an alternative classification, of the 439 public, non-Research I colleges in 1987 that are classified as "Research II," "Doctorate-Granting," "Comprehensive," or "Liberal Arts," roughly 43 percent started as state normal schools. In more recent years, and in the relevant year for our sample of 1978-1983 birth cohorts, most of the previous normal schools were not classified as research universities. Based on the 2000 Carnegie rating, only 15% of the previous normal schools were classified as "Doctoral/Research Universities". Roughly 71% were classified as "Master's Colleges and Universities" and roughly 13% were classified as "Baccalaureate Colleges".

census of "institutions of mental disease" (Furbush et al., 1926). As in Howard, Weinstein and Yang (2024), we focus on institutions established around the same time, so we exclude five asylums that were established before 1830. Counties that had both normal schools and asylums are defined as normal school counties (there are 26 of these counties and in Howard, Weinstein and Yang (2024), excluding these counties had little effect on the results). Our sample includes 204 normal school counties and 125 asylum counties. Figure A1c shows the geographic distribution of normal school and asylum counties in our sample.

1.3 Historical Measures of Mobility

Our identification assumption is that asylum counties are a good counterfactual for what would have happened in normal school counties, had the normal school counties been given a slightly different type of institution. To support this assumption, Howard, Weinstein and Yang (2024) showed balance between normal school and asylum counties in 1840, before most of the normal schools and asylums were established.[13] We add to that support by showing balance on economic mobility in 1850, a year in which we can construct good measures of mobility but that is still before most normal schools were established. In addition, we look at differences in economic mobility for children in 1920—a time in which normal schools had converted to teachers colleges but before they had increased dramatically in size relative to county population and before they were fully converted to regional universities—in order to better understand the mechanisms underlying our findings using recent data.[14]

[13]Besides social mobility, which we discuss here, we also investigate other county-level differences in 1850 in Appendix Table A1, including a summary measure of county-level economic development based on principal components analysis. If anything, normal school counties are smaller in population in 1850 (though not significantly), and there is some evidence they are slightly less urban, with a greater fraction employed as farmers, and have lower real estate wealth per capita in 1850.

[14]Dunham (1969) states that at institutions which train people to be teachers, the students are from lower-middle-income families and often first-generation college students. Ogren (2003) also discusses the normal schools enrolling students from lower-income backgrounds.

1.3.1 Balance on Social Mobility Measures, 1850-1860

First, we construct several measures of economic and social mobility for children in 1850, before most normal schools were established. In this section, we give an overview of our methodology and the main results; see Appendix B for more details. We focus on 16-18 year-old children in the 1850 full count census who are living with at least one of their parents in normal school or asylum counties (Ruggles et al., 2021*a*). We then link these individuals to their record in the 1860 full count census data using the 1850 to 1860 Census Tree crosswalk (Price et al., 2023*a*) and the 1860 full count census data (Ruggles et al., 2021*a*), focusing on 1860 to avoid the effects of the Civil War. Approximately 60% of White males and 40% of White females in between the ages of 16 and 18 living in normal school and asylum counties in our sample have links between their 1850 and 1860 census records. For our primary analysis, we focus on children with low parental socioeconomic status, which we define to be real estate wealth in 1850 less than or equal to 150 dollars.[15]

Table 1 shows the differences in various measures of mobility between people in our sample who were children in normal school and asylum counties in 1850. We estimate the following regression, which we will use throughout the paper:

$$y_i = \beta \text{Normal}_i + \alpha_s + \epsilon_i \tag{1}$$

where y is our outcome of interest from Chetty et al. (2018), i is a county, and α_s is a state fixed effect. The sample consists of counties that had an insane asylum or normal school, and Normal_i is equal to 1 if the county had a normal school. β can be interpreted as an average effect of having been assigned a normal school on the outcome y. As we describe in Appendix B.1, this exercise includes 15 states, and so we do not cluster at the state level given the small number of clusters. Instead, we present standard errors robust to heteroskedasticity

[15] Real estate wealth of $150 is the 37th percentile of the parental real estate wealth distribution in the states in our sample, while wealth of zero is the 33rd percentile. We use $150 to increase sample sizes for these low socioeconomic status children. Appendix B.2 shows that student attendance is increasing in parents' real estate value, suggesting that this measure is meaningful.

Table 1: Economic and Social Mobility for Children in 1850

	(1)	(2)	(3)	(4)	(5)	(6)
	Enrolled	Enrolled	Top Quartile Real Estate	Top Quartile Personal Estate	Married	Occ. Mobility
Year of Observation	1850	1850	1860	1860	1860	1860
Age in 1850	7-13	14-17	16-18	16-18	16-18	16-18
Grew up in Normal School County	-0.032	-0.005	0.008	0.005	0.005	-0.034*
	(.044)	(.037)	(0.016)	(0.013)	(0.016)	(0.017)
Observations	102	102	102	102	102	102
R-Squared	0.507	0.484	0.510	0.548	0.404	0.436
Mean DV, Asylum Counties	0.486	0.378	0.234	0.203	0.593	0.501
p-value randomization inference	0.474	0.899	0.58	0.725	0.761	0.091
Parental SES	Low	Low	Low	Low	Low	All

Notes: $^{+}$ $p < 0.1$, * $p < .05$, ** $p < .01$. Low parental SES (socioeconomic status) is defined as parental real estate wealth in 1850 less than or equal to 150 dollars. Outcomes in columns 1-2 are county-level averages for children living with at least one parent in 1850. Outcomes in columns 3-5 are county-level averages of 1860 outcomes for individuals who were 16-18 year old in 1850 and living with at least one parent in 1850. Occupational mobility in column 6 is the county-level fraction of males with different occupations in 1860 than their father in 1850, among males 16-18 years-old in 1850. Averages in columns 3-6 are calculated among individuals who could be matched to their 1860 records using The Census Tree (Price et al., 2023*a*). We use the 1850-1990 county crosswalk from Eckert et al. (2020). There are 15 states in the regressions. Top quartile real estate and personal estate are based on the distribution of White 25-28 year-olds in our sample of states. Robust standard errors are in parentheses. P-values based on randomization inference are obtained by permuting within states the assignment of normal school counties, among our sample of normal school and asylum counties. This is based on 1000 permutations. See Appendix B.1 for details.

and additionally present p-values based on randomization inference, permuting within states the assignment of normal school counties, among our sample of normal school and asylum counties.

For children growing up in lower socioeconomic status families, there is no difference in the likelihood of having top-quartile real estate or personal estate values or marriage. We see some evidence that occupational mobility—which measures the share of sons in a different occupation than their father—may actually be lower for children growing up in normal school counties. Of course, occupational mobility is not the same as social mobility, but is likely correlated (see Appendix B.1). We show many robustness checks—including principal component analysis, alternative mobility measures based on occupational income scores, and reweighting the Census Tree linked data for representativeness—in Appendix B.1.

While we think that linking children to their future census records provides the closest measure of mobility to our main outcome measures, we also check other measures of mobility that have been used in the literature. In particular, we follow Card, Domnisoru and Taylor (2022) and Derenoncourt (2022) and construct an education-based measure–the likelihood that children of parents with lower incomes or education levels have more education.[16]

We measure upward educational mobility as the school attendance rate among children living with parents of lower socioeconomic status. We use the same measure of socioeconomic status as above—children in families with parental real estate less than or equal to 150 dollars. Columns 1 and 2 of Table 1 show there are no significant differences in upward educational mobility in 1850.

1.3.2 Effects on Social Mobility, 1920-1940

Using a similar methodology as the previous section, we also construct several measures of mobility after the establishment of normal schools, when most offered at least a bachelor's in education but had not yet converted to universities. This is helpful for understanding when these education institutions started conferring mobility benefits on the children living nearby.[17] First, we use the Census Tree to link children in the 1920 census to their records in 1940. We focus on 6-15 year-old children in the 1920 full count census who are living in normal school or asylum counties with at least one of their parents (Ruggles et al., 2021*a*). We then link these records to their record in the 1940 full count census data using the 1920 to 1940 Census Tree crosswalk (Price et al., 2023*b*) and the 1940 full count census

[16] Derenoncourt (2022) uses the occupational score of the fathers to identify socioeconomic status, but this is based on 1950 incomes and this score could be quite different in 1850. Specifically, among seven to seventeen year-old children in 1850 who were living with their fathers, 60 percent had fathers who were farmers, and 85 percent had fathers whose occupation was in one of five codes (farmer, manager, carpenter, laborer, operative). Card, Domnisoru and Taylor (2022) uses the educational attainment of the parents, but this is not available in the 1850 census.

[17] We also look at other economic covariates in 1940 in Appendix C.1, including summary measures based on principal components analysis. While there are some differences between the counties, they are small, and do not show a major impact of the normal schools on the local economy, at least outside of education and social mobility which we discuss in this section. Appendix B.5 shows an alternative measure of mobility for 1920-1940 using conditional mean occupational income score ranks.

data (Ruggles et al., 2021*a*). Details on the methodology, results, and additional findings, including reweighting the Census Tree linked data for representativeness, are in Appendix B.3.

We do not see statistically significant differences in occupational mobility (Tables A10 and A11). However, we do find evidence that among lower socioeconomic status children, the fraction with some college and college completion is higher among those who grew up in normal school counties relative to those who grew up in asylum counties (Table 2). The magnitudes are larger for women, who are also more likely to be married and less likely to be employed in 1940 if they grew up in normal school counties. There are no statistically significant differences in household wage and salary income, which would also include income from work relief.[18] We additionally see effects of growing up in a normal school county on college attainment, for men and women from higher socioeconomic status families (Table A15).

These results suggest the teachers colleges and the state colleges that the normal schools had evolved into by the 1930s were already affecting access to education in their local communities. Our analysis in the next section tests whether these effects continue in recent years, when there were dramatic increases in the colleges' size but also potentially decreases in geographic frictions in college attendance.

Table A38 shows there are no significant differences in the measures of upward educational mobility in 1940 from Card, Domnisoru and Taylor (2022)–fraction of children attaining 8th and 9th grades among those living with parents who completed no more than 6th grade.

1.4 Effect on Higher Education and the Economy, 1980

In this section, we show the impact of normal school assignment on higher education in 1980, around the time the Opportunity Insights sample was born. These results are discussed further in Appendix A.1.

[18]Table A14 shows differences in likelihood of work relief between normal school and asylum counties.

Table 2: Economic and Social Mobility for White Children in 1920, Measured in 1940

	(1)	(2)	(3)	(4)	(5)	(6)
	$\geq$ HS	$\geq$ Some College	$\geq$ College	Married	Employed	Ln(HH Income)
Panel A: Males, Lower Socioeconomic Status						
Grew up in Normal School County	0.007	0.007^{+}	0.002	0.005	-0.006^{+}	-0.026
	(0.007)	(0.004)	(0.002)	(0.004)	(0.003)	(0.016)
Observations	314	314	314	314	314	314
R-Squared	0.583	0.497	0.425	0.640	0.362	0.495
Mean DV, Asylum Counties	0.308	0.113	0.055	0.707	0.901	7.027
Panel B: Females, Lower Socioeconomic Status						
Grew up in Normal School County	0.010	0.024**	0.008**	0.019*	-0.020**	-0.013
	(0.010)	(0.006)	(0.003)	(0.008)	(0.007)	(0.015)
Observations	314	314	314	314	314	314
R-Squared	0.474	0.408	0.355	0.675	0.657	0.471
Mean DV, Asylum Counties	0.404	0.140	0.047	0.701	0.297	7.052

Notes: $^{+}$ $p < 0.1$, * $p < .05$, ** $p < .01$. Outcomes are county-level averages, among individuals who could be matched to their 1940 records using The Census Tree (Price et al., 2023*b*). Lower socioeconomic status individuals are defined as individuals whose parents' maximal occupation score in 1920 was less than or equal to the median for the states in our regression sample (the median score was 20). The variable Married is the fraction of individuals who are married with spouse present. The variable Ln(HH Income) is the log of the county-level average wage and salary income of individuals and their spouses. The county-level average is constructed by taking the total wage and salary income of individuals and their spouses and dividing by the total number of individuals (males in Panel A and females in Panel B) with positive own or spousal wage and salary income. We use the 1920-1990 county crosswalk from Eckert et al. (2020). Standard errors are clustered at the 1920 state level.

First we show that normal school counties are more likely to have a large, public university relative to asylum counties, which we also showed in Howard, Weinstein and Yang (2024). In 1980, 91 percent of counties that were historically assigned a normal school have a regional public college or university that had been a normal school, while this percentage is mechanically zero in asylum counties (Appendix Table A2).[19] Some asylum counties have public four-year colleges, and the within-state differences imply normal school counties have 0.7 additional public four-year colleges than asylum counties. Because not all normal school counties have a regional public college, and some asylum counties do have a public four-year college, this implies that our reduced-form empirical strategy will underestimate the impact of regional public universities.

On average, asylum counties have more private four-year colleges and two-year colleges. Adding the coefficients for total public four-year, private four-year, and two-year colleges suggests a similar number of colleges in the two types of counties. However, the universities in the normal school counties are much larger. Enrollment as a percent of population is an additional 8.5 percentage points higher in normal school counties, with enrollment equal to 4.6% of population in asylum counties. Finally, the fraction of the population with a college degree is 2 percentage points higher in normal school counties, which is large relative to the level, though small relative to the number of degrees awarded per year as a percent of population. This suggests many students leave after graduating.

Table A3 further shows that the net price of attending institutions in normal school counties is lower than institutions in asylum counties, bachelor's completion rates are lower, and test scores of incoming students are also lower.

We also wish to emphasize economic comparisons between normal school and asylum counties. As stressed in Howard, Weinstein and Yang (2024), the 1980 economies in normal school and asylum counties look similar in both levels and growth, with the biggest difference being that normal school counties have a slightly larger retail sector and a slightly smaller

[19]For an additional two counties, the normal school closed and the site of the normal school became a different university. This was true of UCLA and Maine Maritime Academy.

manufacturing sector.

Given the focus of our paper on social mobility, we also wish to highlight the income distribution of normal school and asylum counties. Figure 1 shows that the fraction of parents in each national income decile is similar in normal school and asylum counties within the same state. The specification for the regression shown in Figure 1 is the same as our main specification, equation (1). Normal school counties have slightly higher percentages in the fourth thorough sixth deciles, and lower percentages in the ninth and tenth. Figure 1 also makes clear that asylum and normal school counties are different than the country as a whole, with substantial underrepresentation of people with very low incomes, as well as underrepresentation of people with the highest incomes. If counties were representative of the country as a whole, then 10 percent of the population would be in each decile.

2 Effects on Local Social Mobility, 2005-2015

2.1 Data on Economic and Social Mobility

For our primary outcomes, we obtain data from Chetty et al. (2018). Using IRS and census data, this includes county-level outcomes of children born between 1978 and 1983 who grew up in the county, by their parents' income. The sample includes 96 percent of all children born between 1978 and 1983, who were born in the U.S. or are authorized immigrants who arrived in the U.S. as children and whose parents were U.S. citizens or authorized immigrants. Parents are defined as the first person who claims the child as a dependent between 1994 and 2015. Individuals are attributed to a county in Chetty et al. (2018), weighted by the fraction of years that they are claimed as a dependent in the county before age 23. Importantly, these estimates of child outcomes in adulthood at the county level are based on the children who grew up in the county, regardless of whether they still live in the county when outcomes are measured in adulthood.

We test for effects on educational attainment, income and employment, and other social

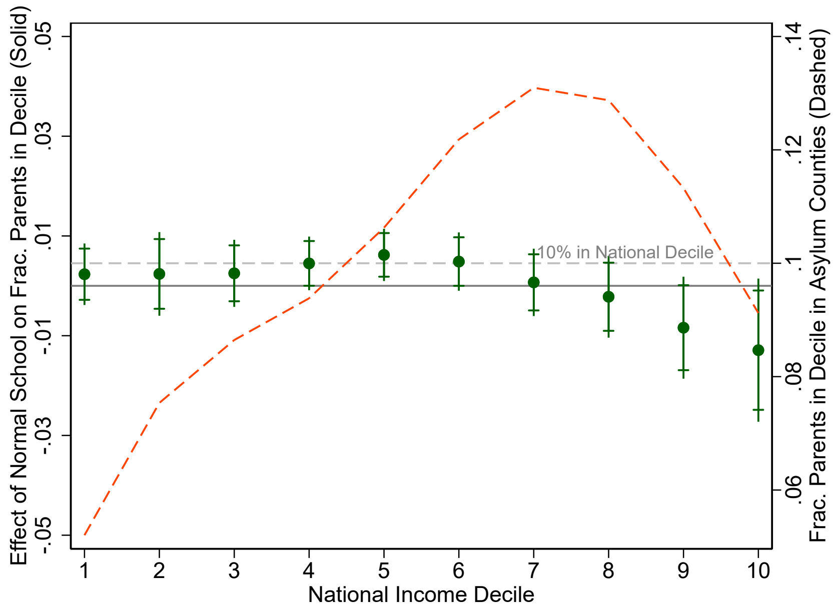


Figure 1: **Fraction of Parents by National Income Decile.** Green dots with spikes represent the estimated effect of a normal school on the outcome variable, conditional on state fixed effects. Outcome variables are the fraction of parents in the county in each national income decile. We estimate a separate regression for each decile, with effects across the x-axis. Standard errors are clustered at the state level. The green spikes span the 95 percent confidence intervals, with crossbars at the 90 percent confidence intervals. The estimates and confidence intervals correspond to the y-axis on the left-hand side of the figure. The orange dashed line shows the mean of the outcome variable in asylum counties, and corresponds to the y-axis on the right-hand side of the figure. Fraction of parents in national income decile, from Chetty and Hendren (2018), is based on parents of children in the 1980-1986 birth cohorts and average family income over 1996-2000.

outcomes. For education, we analyze the fraction obtaining at least a four-year degree, the fraction with some college, and the fraction with at least a high school degree or a GED. These education outcomes are observed only in the ACS, and thus are only available for the subsample that is observed in the ACS between 2005 and 2015. The number of children in this subsample is roughly four million, relative to the full sample of 20.5 million. The fraction obtaining at least a four-year degree, and the fraction with some college, are measured only for people at least 25.[20] The fraction with at least a high school degree or GED is measured only for those at least 19.

[20]Median age at graduation was 23 for public four-year institutions that were not very high research activity based on the 2005 Carnegie ratings (U.S. Department of Education, National Center for Education Statistics, 2021). Thus, this sample restriction will likely not capture too many people who are still enrolled and have yet to obtain a degree.

The income and employment outcomes we analyze include the fraction with positive W-2 earnings in 2015, family income percentile in 2014-2015, and individual income percentile in 2014-2015. Children's income as an adult is measured as the average of their adjusted gross incomes in 2014 and 2015, when they are 31-37 years old. The other social outcomes we analyze include the fraction married in 2015, teen birth (for women only), fraction incarcerated on April 1, 2010, fraction staying with their parents in 2015, and fraction staying in their childhood commuting zone based on their most recent address. These income, employment, and social outcomes are observed for the full sample.[21]

Chetty et al. (2018) provide predicted children's outcomes in each county at five different percentiles of the parental income distribution.[22] Parental income is measured as the mean of parents' household adjusted gross income in 1994, 1995, and 1998-2000, when children are 11-22 years old. Given the children's age when parents' income is measured, we are less concerned that lower-income children in normal school counties are the children of graduate students, who may be experiencing only temporarily reduced income levels.

2.2 Empirical Strategy

Throughout this section, we continue to estimate equation (1), comparing same-state counties that were assigned normal schools and asylums. The identification assumption is that asylum counties in the same state are a good counterfactual for the social mobility of normal school counties, had the normal school county been assigned a different type of institution

[21] Additional variables details are described in Appendix D.1. The measures in Chetty et al. (2018) that we do not use as outcomes in our analysis were mostly either non-baseline definitions of variables, or employment and income measures from the ACS (we use the IRS measures given the larger sample). The other child outcomes in adulthood that we do not include are the fraction with a community college degree (we use some college instead), and the fraction with a graduate degree. We also did not include the fraction living in low-poverty tracts as adults, or in their same census tract (we use fraction remaining in their CZ and fraction remaining with their parents).

[22] These predictions are based on regressing children's outcomes on parents' income percentiles, and allowing the coefficient to vary by county. Chetty et al. (2018) parameterize the relationship between child and parent income using a lowess regression of children's outcomes on parent's income percentile at the national level.

instead.[23] We test many outcomes in this section, and address questions about multiple hypothesis testing in Appendix D.2.

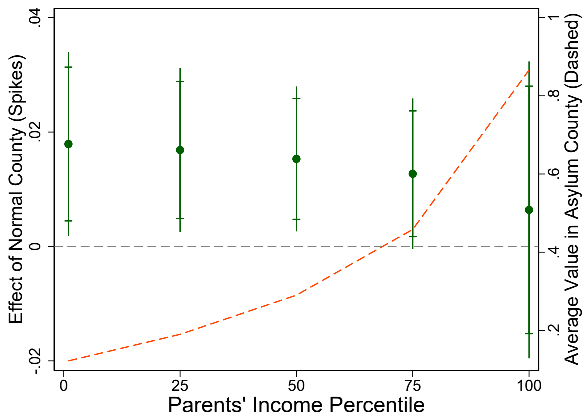
2.3 Effects on Education

We first study the educational attainment of children who grow up in the county.

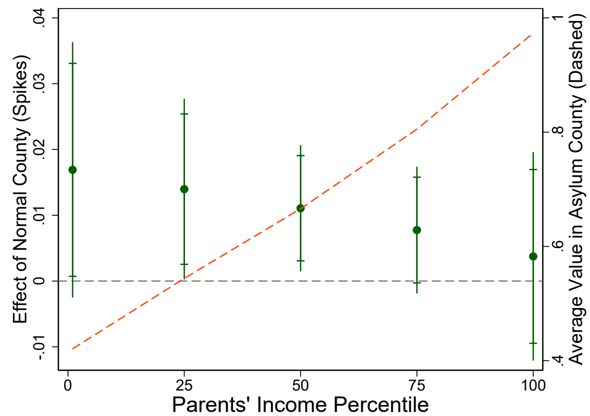
The regression results are shown in Figure 2. The green dots are the estimated coefficients from regression (1). The spikes are the 95 percent confidence intervals, and the cross-bars are 90 percent confidence intervals. The x-axis is the parents' income percentile, so the estimates to the right are for children of high-income parents, and the estimates to the left are for children of low-income parents. The estimates correspond to the y-axis on the left. For example, in panel (a), the effect of having been assigned a normal school is about a 1.8 percentage point increase in the probability of getting a four-year college degree, for a child who grows up in that county with a parent at the 1st percentile of the national income distribution. In the dotted orange line, the mean value of the outcome in asylum counties is plotted against the parents' income percentile, and the corresponding y-axis is on the right. For example, at the 1st percentile, about 12 percent of the children in asylum counties get a four-year college degree. The orange line is not a causal estimate, but provides important context for interpreting the magnitudes of the effect. Note that the scales on each axis are different and vary from figure to figure.

In panel (a), we see a significant increase in college degree attainment for children growing up in normal school counties, by between one and two percentage points for children of parents at the 1st, 25th, 50th, and 75th percentiles. For the 100th percentile, the point estimate is smaller and the confidence interval is quite wide. While there are not significantly

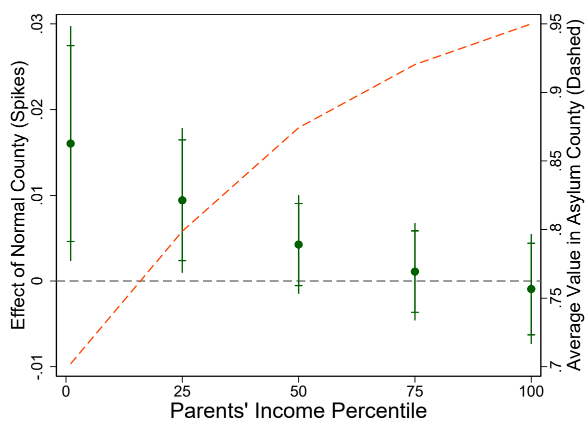
[23] Throughout the section, we view our hypothesis test as testing whether the direct effects of a university on its own county are different than its effects on same-state asylum counties. It would also be interesting to see if there are differential spillovers on nearby asylum counties compared to far-away asylum counties. In Appendix D.3, we investigate this question by splitting our sample of asylum counties based on their distance to a normal county. However, that test is underpowered. It can neither reject that the spillovers to close asylum counties are as large as the direct effect on normal counties nor reject that the spillovers to close asylum counties are the same as those on the far-away asylum counties.



(a) At least 4-year College Degree, Age ≥ 25



(b) At Least Some College, Age ≥ 25



(c) At least HS Graduate or GED, Age ≥ 19

Figure 2: **Effect of a normal school on local children's education, 2005-2015.** Green dots with spikes represent the estimated effect of a normal school on the outcome variable, conditional on state fixed effects. Outcome variables are measured at different percentiles of the parents' income distribution, which is the x-axis. The green spikes span the 95 percent confidence intervals, with the cross-bars at the 90 percent confidence intervals. Standard errors are clustered by state. The estimates and confidence intervals correspond to the y-axis on the left-hand side of the figure. The orange dashed line shows the mean of the outcome variable in asylum counties, and corresponds to the y-axis on the right-hand side of the figure.

significant differences in the effects across the income distribution, the effects for lower income percentiles are much larger relative to the baseline. For a child with parents at the 75th percentile, the increase is less than 3 percent of the baseline, while it is about a 15 percent increase for a child at the 1st percentile, and 9 percent for a child at the 25th percentile.

In Panel (b), we look at the effects on some college attendance. The point estimates are generally comparable, which is noteworthy for two reasons. First, if proximity to normal schools only affected substitution between two- and four-year colleges, or only affected completion among those who enroll, then the effects on some college would be zero. While the effect of normal schools on four-year attainment does not appear driven by either of these explanations, they do seem more evident among students as parental income rises. Second, if normal schools increased enrollment in four-year colleges, but these marginal students were unlikely to complete a degree, the effects on some college would be larger than the four-year effects. However, that is not what we find.

In Panel (c), we find significant increases in the high-school degree or GED attainment for children with parents at the low-end of the income distribution. The point estimate at the 1st percentile is comparable to the point estimate of the effect on some college or the effect on four-year college degree attainment. If normal schools' only effect was incentivizing high school graduates to enroll in college, there would be no effect on high school completion.[24]

We note our results are in contrast to Bedard (2001), who finds that proximity to a college increases the high school dropout rate among teenagers in the 1960s. A central difference in our analyses is the identification of the control group.

The results from our causal identification strategy confirm the results of Card (1993) and the subsequent literature, that has used proximity to a college as a predictor of college attendance. For comparison, Kling (2001) shows that for the lowest-quartile of family

[24] The magnitudes we find on high school completion are suggestive that there may be effects of the university on the quality of local K-12 education, since the magnitudes for college graduation and high school graduation are similar for children from low-income families. In fact, in Appendix C.3, we see that normal school counties have lower student-teacher ratios, which could be due to the fact that regional universities educate many future teachers. The higher high school graduation rate may also be due to increased incentive to graduate high school to attend college, due to lower geographic frictions.

background, having a college in the county increases highest grade completed by roughly one year in 1976 for individuals who were 14-19 in 1966. In 1989, this had fallen to 0.5 years for individuals who were 14-19 in 1979. While not directly comparable to our outcome variables, our empirical strategy appears to yield substantially smaller effects.[25] One reason may be that colleges are located in areas that have higher attainment for reasons other than the college, and our empirical strategy accounts for those. In our strategy, colleges may affect attainment through the direct effect on students and also through indirect effects (e.g., on the economy), but we eliminate the non-causal relationship between colleges and local educational attainment.

Russell, Yu and Andrews (2022) finds a substantially larger effect on college attainment. Their baseline estimate, using data from the American Community Survey, is that the presence of a university increases the share of the population with a college degree by 14 percentage points. When using data from Chetty et al. (2018), they estimate the fraction with at least a bachelor's degree is 8 percentage points higher for people who grew up in counties with universities, which is still substantially larger than our estimate. This likely reflects different effects of top-tier flagships and private universities on their local economies, compared to the effects of regional public universities which are our focus. The universities in Russell, Yu and Andrews (2022) have larger effects on the local industry composition than the universities in our sample (see Howard, Weinstein and Yang, 2024). That could mean the research-intensive universities attract parents more likely to send their children to college relative to the regional public universities in our sample. It could also be that the universities in Russell, Yu and Andrews (2022) provide a higher return to a college degree.

In Appendix D.6, we look at the effects on education by race and sex. The sample of

[25] The reason this comparison is challenging is that we do not observe years of education, the main outcome variable in those studies. However, if both point estimates are correct, it would need to be the case that almost all of the increase in years of schooling is due to students who do not obtain an additional degree. For example, if we take our biggest coefficients for each outcome, and assume that every additional college graduate or high school graduate gets another four years of schooling, that would contribute only 0.14 additional years of schooling ($.018 \times 4 + .016 \times 4$), which means that 0.36 years would have to come from students who get more schooling but not additional degrees.

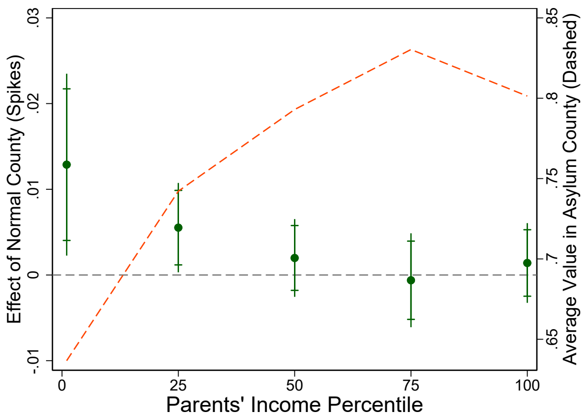
counties is different across races due to data availability, making comparisons across race difficult. The results are also noisier, making it hard to say anything conclusive. However, there are several interesting observations within race. For Hispanics, the effects on high school attainment are very large for those from lower-income families. And for some of the results regarding college degrees and some college, the effects for black and Hispanic children are the strongest at the top of the income distribution. For college degrees, the effects are stronger for women, while for high school degrees, the effects are slightly stronger for men at the bottom of the income distribution relative to women.

2.4 Effects on Income

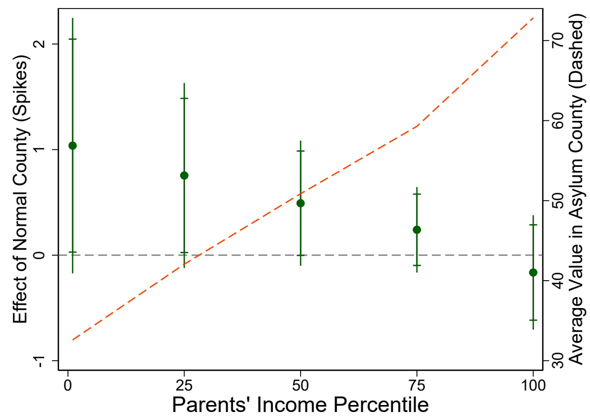
In Figure 3, we show effects on measures of income from Chetty et al. (2018). Panel (a) shows the effect on having any positive wage income in 2015, when the sample is 32 to 37 years old. At the first percentile of the parental income distribution, regional public universities increase the probability of positive W-2 earnings by 1.3 percentage points, which is an increase of 2 percent relative to the baseline. At the 25th percentile of parental income, there is a 0.6 percentage point increase, which is an increase of 0.7 percent. Recall that we see a 1.7 percentage point increase in four-year degree attainment, at the 25th percentile of parental income. If the 0.6 percentage point increase in employment is driven by the 1.7 percentage point increase in education, this implies large positive employment effects on the additional degree recipients.

When we look at the family income percentile or the individual income percentile, there is an increase that is more pronounced at the low-end of the distribution and the family income results are borderline significant at conventional levels. We find that regional public universities raise household income percentile rank of children at the 25th percentile by roughly 0.8 percentile ranks (p-value ≤ 0.1), when measuring their incomes in 2014-2015 at age 31-37.[26]

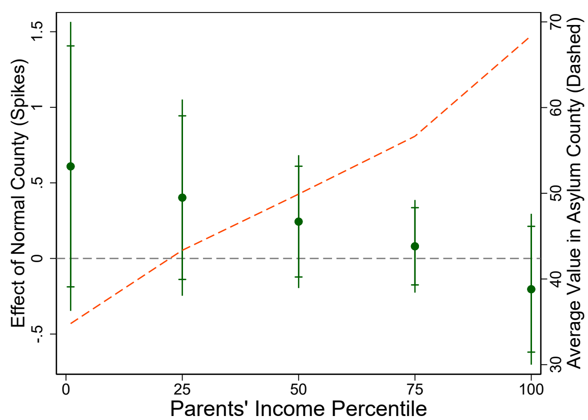
[26]For comparison, Chetty and Hendren (2018) show that growing up in a commuting zone with one



(a) Positive W-2 Earnings, 2015



(b) Family Income Percentile, 2014-2015



(c) Individual Income Percentile, 2014-2015

Figure 3: **Effect of a normal school on local children's adult income, 2014-2015.** Green dots with spikes represent the estimated effect of a normal school on the outcome variable, conditional on state fixed effects. Outcome variables are measured at different percentiles of the parents' income distribution, which is the x-axis. The green spikes span the 95 percent confidence intervals, with the cross-bars at the 90 percent confidence intervals. Standard errors are clustered by state. The estimates and confidence intervals correspond to the y-axis on the left-hand side of the figure. The orange dashed line shows the mean of the outcome variable in asylum counties, and correspond to the y-axis on the right-hand side of the figure.

Given that a likely mechanism for the increase in income is the increase in college-going, it is worth comparing our results to measures of the return to college education. At the 25th percentile of parental income, our results imply a roughly 2 percent increase in income for children who grow up in normal counties.[27] Given that we estimated a slightly smaller than 2 percentage points increase in college attainment, this would require the returns to a college degree of about 100 percent. Using cutoffs for entrance to a less-selective college—like the ones studied here—Zimmerman (2014) estimates a return of 22 percent to college admission for the marginal student 8-14 years after high school completion. If college completion were the only mediator, he finds the return to a college degree would be 90 percent.[28] While an admittedly rough comparison, this means that the increase in college completion can explain a large fraction of the effect that we find on income. However, it may not be the whole story given the effects we also find on high school completion.

To put it in comparison to the baseline, the slope of the social mobility curve in asylum counties (the orange dotted line) is about 0.4. Taking the point-estimates at face value, the causal effect of being assigned a normal school would be to reduce that by about 0.012, or about 3 percent. Both the slope and the impact are somewhat muted when focusing on individual income, with normal schools reducing the slope by about 2.4 percent.

While the confidence intervals are large, it is of note that the effects on college attainment

standard deviation lower racial segregation is associated with higher household income rank of children at the 25th percentile by 1.6 percentile ranks. One standard deviation lower income segregation is associated with higher rank by 1.1 percentile ranks. That is a purely correlational result, while the 0.8 percentile rank increase we identify is the causal effect of regional public universities on local children.

[27]Using a different set of birth cohorts, and measuring income at a different age than in our sample, Chetty and Hendren (2018) show that for the 1980-1986 birth cohorts, an increase of one percentile rank in household income at age 26 translates to an additional 818 dollars, for children whose parents were at the 25th income percentile, which is an increase in income of roughly 3.14 percent. If the relationship between percentile rank and percent increase in income holds for the slightly older individuals in our sample, our results would imply regional universities increase income by roughly 2.4 percent for children who grew up in the county with parents at the 25th income percentile.

[28]Other papers find estimates of similar magnitudes: Card (1993) estimates returns of about 13 percent per year of education using an instrumental variable strategy based on proximity to college, Smith, Goodman and Hurwitz (2020) estimates returns to enrollment in a Georgia public four-year college of 20 percent, and Kozakowski (2023) estimates returns to admission to four-year public colleges in Massachusetts for low-income students of 26 percent. All of these variables would have to be scaled up significantly to be about college completion rather than college admission or enrollment.

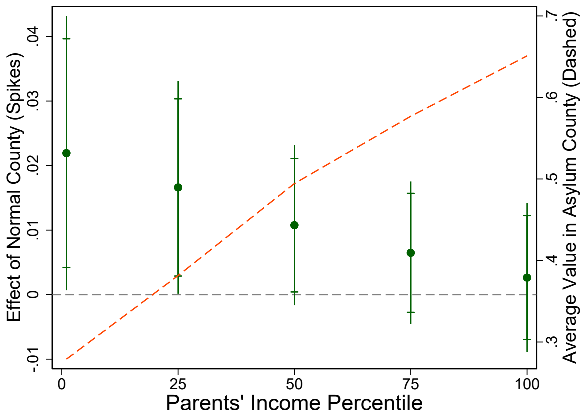
were roughly constant across parental income, but the effects on employment and income are much more pronounced for children from lower-income families. This is consistent with the additional enrollees experiencing stronger labor market benefits of college if they were from lower-income families.

Russell and Andrews (2022) looks at the effects of universities on income rank, although they estimate the effect of primarily research-intensive universities. For children born to parents at the 1st or 25th percentile, they find an increase in the mean income rank in 2014-15 of 0.003, although the effect is insignificant. This is somewhat smaller than our estimated effect of about 0.01. Interestingly, given our smaller effect on education attainment, these results could suggest a higher income return to regional universities, although there are certainly other possible mechanisms, and the estimated effects are not particularly precise for either type of university.

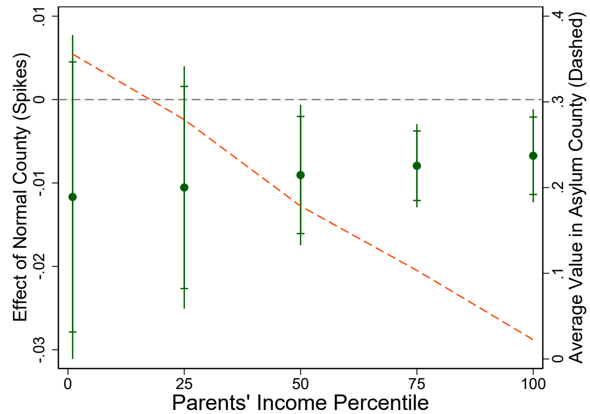
2.5 Effects on Other Social Outcomes

We also examine the effects of being assigned a normal school on marriage, teen childbirth, incarceration, and migration. These results are presented in Figure 4. In Panel (a), we look at the effects of normal school assignment on marriage rates across the parental income distribution. Consistent with the larger effects we found on family income relative to individual income, we find positive effects on marriage in 2015 when the sample is age 32 to 37, with children born to parents in the 1st percentile being 2.2 percentage points more likely to get married, approximately a 8 percent increase. For the 25th percentile, the increase is 1.7 percentage points, roughly a 4 percent increase.

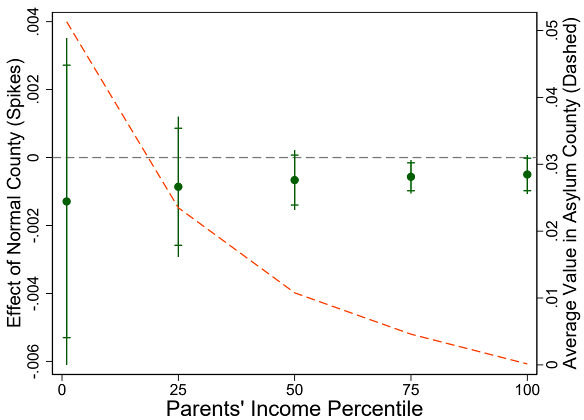
In Panel (b) we also find negative effects on teen childbirth. The point estimates are larger for children of lower-income parents, but the standard errors are also larger, so the only statistically significant results are at the top half of the distribution. However, these are large: about 1 percentage point across the distribution, off of a baseline ranging from 2 percent at the top end to about 36 percent for children of the lowest-income parents.



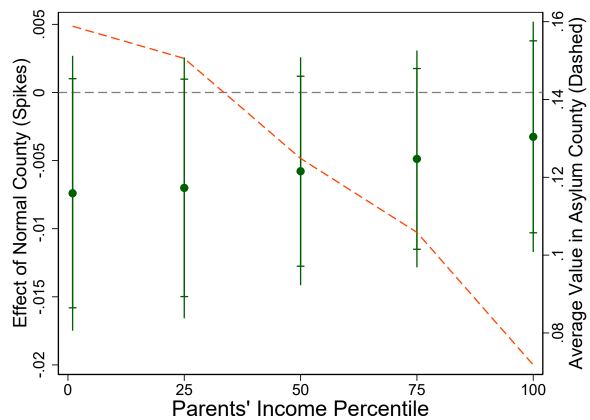
(a) Marriage, 2015



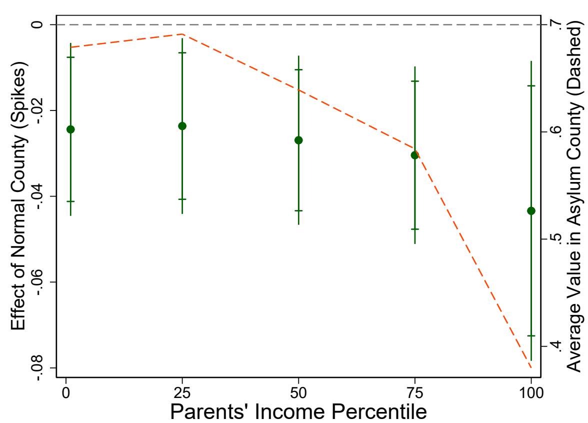
(b) Teen Birth (women-only)



(c) Incarcerated, April 1, 2010



(d) Staying with Parents, 2015



(e) Live in Childhood Commuting Zone

Figure 4: **Effect of a normal school on local children's social outcomes.** Green dots with spikes represent the estimated effect of a normal school on the outcome variable, conditional on state fixed effects. Outcome variables are measured at different percentiles of the parents' income distribution, which is the x-axis. The green spikes span the 95 percent confidence intervals, with the cross-bars at the 90 percent confidence intervals. Standard errors are clustered by state. The estimates and confidence intervals correspond to the left y-axis. The orange dashed line shows the mean of the outcome variable in asylum counties, and correspond to the right y-axis.

Panel (c) shows negative effects on incarceration. As with teen birth the point estimates are larger for children of lower-income parents, but the results are more precise for children of higher-income parents. For example, for children whose parents were at the 75th percentile, regional public universities reduce the fraction that were incarcerated on April 1, 2010 by 0.05 percentage points, from a baseline rate of 0.5 percent in asylum counties.

Panel (d) shows that children are less likely to live with their parents in 2015 if they grew up in a county that had been assigned a normal school. However, these results are not statistically significant.

Panel (e) shows that in normal school counties, children are less likely to remain in the commuting zone in which they grew up. The effect is more negative for children of high-income parents, despite already having a much lower baseline. These effects are large, with children in normal school counties being about 3 percentage points more likely to move out of the commuting zone, and about 4.3 percentage points for the 100th percentile of parental income.[29] Taking together these results with our results on economic mobility, and the absence of impacts on the local economy, suggests proximity reduces frictions in college attendance and this raises mobility through increasing geographic mobility.

2.6 Comparison to Causal Effects on People

Our estimates in Figures 2, 3, and 4 identify the causal effects of having a been assigned a normal school on county characteristics. While these estimates identify the causal impact on the place, they do not identify the causal impact on an individual child, because the university may also affect the composition of children who grow up in the county.[30] We use Chetty and Hendren (2018) estimates of causal effects of an additional year of exposure to

[29] We see larger effects for children from the 100th percentile, even though there were not statistically significant effects on educational attainment for this group. This may reflect that children of faculty and higher-level university administrators are more geographically mobile, given the likely greater geographic mobility of their parents. In asylum counties, it is less likely that the higher-income families are faculty or university administrators. As we show in Table A39, there is some evidence that children in normal school counties spent slightly less of their childhood in the commuting zone than children in asylum counties.

[30] Chetty and Hendren (2018) show evidence consistent with lower-mobility, not higher-mobility, individuals sorting into areas with more colleges per capita.

a county, which accounts for this selective location choice, to see if the college does indeed have an effect on the outcomes of a child.[31]

In Table 3, we use our same empirical strategy but use the causal estimates on individuals from Chetty and Hendren (2018) as the dependent variables. We focus on students with parental income at the 25th percentile, but show the 75th percentile in Appendix D.5. We use the outcome that is most comparable between the two datasets. Columns (1), (3), and (5) show the same results as in Figures 2b, 3b, and 4a, for the 25th percentile, the effects on some college attendance, family income percentile, and marriage, respectively. Columns (2), (4), and (6) look at the closest variable in the Chetty and Hendren (2018) dataset, although the precise definitions are slightly different for each variable. Unfortunately, other outcomes are unavailable in the Chetty and Hendren (2018) dataset.

There are a few differences to note when comparing these columns. First, following Chetty and Hendren (2018), to maximize precision when analyzing the causal impacts on people, we weight the observations using the inverse of the variance of the estimate. These weights are correlated with county population, so if the effect size is correlated to the size of the county, then the coefficients may reflect different average effects. Second, the causal estimates in column (2) are to be interpreted as the effect of having one additional year in that county, whereas the scale in column (1) is based on a childhood. The suggested comparison would be to scale the coefficient in column (2) by about 15 or 20 (see Derenoncourt (2022) for a discussion). Third, the variables are slightly different. For example, column (1) is from ACS respondents and column (2) is from 1098-T forms that universities file with the IRS. Finally, the results are based on different birth cohorts.[32]

At the 10 percent level, there is evidence that having a normal school has a causal effect on the person, for all three outcomes. If we took the point-estimates seriously, it would

[31] Chetty and Hendren (2018) estimates these causal effects using children that move into a county at different ages. One potential concern would be if the types of families who move into a county with older versus younger children is different in normal school versus asylum counties. We do not find evidence of this in Appendix D.7.

[32] Appendix Table A40 shows the causal estimates without weights and the observational estimates with the same weighting scheme as the causal estimates.

Table 3: Causal Effects on People, 25th percentile parental income

	(1) Some College Age 25+	(2) Attended College Age 18-23	(3) Family Income Percentile, 2014-15	(4) Family Income Percentile, Age 26	(5) Married 2015	(6) Married Age 26
Normal	1.397*	0.143+	0.755+	0.0773+	1.662*	0.0856+
	(0.680)	(0.0764)	(0.434)	(0.0436)	(0.817)	(0.0470)
Observations	324	305	324	305	324	300
Weights	Unweighted	Precision Weights	Unweighted	Precision Weights	Unweighted	Precision Weights
Scale	Per Childhood	Per Year	Per Childhood	Per Year	Per Childhood	Per Year
Interpretation	Effect on Place	Effect on Person	Effect on Place	Effect on Person	Effect on Place	Effect on Person

Standard errors clustered by state. $^{+}$ $p < 0.1$, * $p < .05$, ** $p < .01$. Outcome data in columns 1, 3, and 5 are from Chetty et al. (2018), and outcome data in columns 2, 4, and 6 are from Chetty and Hendren (2018).

seem that the causal effects are a bit bigger for college attendance and family income, but between the large standard errors, the different samples, and the different weightings, our main takeaway is that the magnitudes are roughly similar.

3 Mechanisms: Evidence from The Freshman Survey

In this section, we use The Freshman Survey to understand the channel through which regional public universities raise local college-going. In particular, we will evaluate three potential channels: geographic proximity lowers financial frictions, information frictions, or frictions for those with strong preferences for staying near home.[33]

A related question is why students in asylum counties are not equally served by the universities in their counties. Appendix A.1 shows that the four-year colleges that are in asylum counties are more expensive, more selective, and smaller, and many asylum counties do not have any four-year universities.

The Freshman Survey (TFS) has been conducted since 1966 through the Higher Education Research Institute at the University of California, Los Angeles. The survey is administered by colleges to their freshman classes during orientation and has been administered by over 1900 institutions surveying 15 million students (Higher Education Research Institute,

[33] A number of studies have analyzed information interventions to increase college attendance among low-income high-achieving students (see for example Dynarski et al. (2021) and Andrews, Imberman and Lovenheim (2020)).

2023). The survey asks why the student decided to attend college, why the student chose their university, and includes questions about many individual characteristics. It also records the student's home zip code, which we use to identify students who grew up in normal school and asylum counties.[34] With our sample restrictions, the dataset consists of over 2.1 million individuals growing up in all 204 normal school counties, attending 1466 universities from 1982-2010, and nearly 1.5 million individuals growing up in all 125 asylum counties attending 1431 universities. Additionally restricting to students with positive TFS survey weights yields a sample of 2.5 million individuals. See Appendix E for the details on the sample.

We have three main objectives with TFS data. First, our education results in Figure 2 show students growing up in normal school counties are more likely to attain a college degree. With TFS data, we show whether this is driven by enrollment at the nearby previous normal schools. Second, we show whether students growing up in normal school counties have characteristics of more marginal college-goers, consistent with the greater college enrollment among people growing up in normal school counties in Figure 2. Finally, we look for evidence on the specific types of geographic frictions that matter for attending regional universities, e.g. financial costs, informational frictions, or preferences for living near home.

Before we test for differences in the characteristics of freshman from normal school and asylum counties, we first confirm that the universities in normal school counties are not differentially more likely to participate in TFS relative to universities in asylum counties. This would bias our analysis of whether students growing up in normal school counties are more likely to attend college near home. Details are in Appendix E.1.

We then test whether individuals from normal school counties are more likely to attend university close to where they grew up, relative to individuals from asylum counties in the same state. We estimate specification (1), aggregating up to the county level using the survey

[34] The student zip code data become available starting with the 1982 survey. Through 2000, students were asked for their address at the top of the survey, including their zip code. Starting in 2001, students were asked for their "permanent/home address", including their zip code. We acknowledge we do not have information on how long they may have lived at that address. Appendix E.2 addresses the possibility that students are reporting their address at the university rather than their home address.

Table 4: Differential Likelihood of Attending University Close to Home

	(1) Attend Univ. within 10mi	(2) Attend Former-Normal within 10mi	(3) Attend Univ in county	(4) Attend Former-Normal in county
Grew up in normal school county	0.0462**	0.0787***	0.149***	0.202***
	(0.0131)	(0.0127)	(0.0313)	(0.0297)
Observations	324	324	324	324

Standard errors clustered by state. * p<0.05, ** p<0.01, *** p<0.001
Observations are at the county level. All regressions include state fixed effects.

weights provided by TFS (see Appendix E.1 for details). The regression includes data from individuals responding to TFS who report home zip codes located in normal school or asylum counties.[35]

Our first outcome variable of interest is whether a student attends a university within ten miles of home. In addition, we test whether they are attending any university that is in the same county as their home. We show both measures of proximity because students are asked directly their home-university distance, but we have to infer their county based on the zip code they list on the survey. While the county results are more closely related to our previous findings, the within-10-miles results are more robust to concerns about zip code misreporting. We also show the results where y_i is defined as attending a previous normal school within 10 miles or attending a previous normal school in the same county, in order to show how much these universities drive the results.

Students who grew up in normal school counties are 4.66 percentage points more likely to attend a university within 10 miles of where they grew up relative to students who grew up in asylum counties in the same state (Table 4, column 1). The mean of this dependent variable in asylum counties is 8.9%. More than 100 percent of this effect is driven by students that are attending the former-normal school within 10 miles (column 2), meaning that enrollment at other types of universities is crowded out by the presence of a former normal school. When we consider the alternative measure of whether students are attending a university in their home county, the fraction among students growing up in normal school counties is 15 percentage

[35]In our baseline specification we average over respondents in different survey years. Appendix E.4 shows this aggregation does not lead to meaningfully different results relative to a regression specification that compares responses within the same state and year.

points higher (column 3). Again, this is more than 100 percent driven by students attending the former-normal school (column 4).[36]

To better understand the channels through which regional universities raise local college-going, we estimate equation (1), with various TFS questions as the dependent variables. Many questions ask students to rank something as important on a three point scale. In this section, we create an indicator variable for either "Very important" or "Somewhat important" and use that as the outcome variable.[37] We emphasize that the coefficients β reflect differences in who chooses to attend college (and which college) in normal school versus asylum counties, along with any effects of the regional university on characteristics of children growing up nearby. For many of the regressions, the most natural interpretation is that the regional university is changing the composition of students that attend college, but it is not the only possible interpretation.[38]

Our results are presented in Figure 5. In each figure, we plot the estimate of the coefficient on growing up in a normal school county. Given the many outcomes we look at and that these results are primarily interesting because they help elucidate the mechanisms in the previous sections, we do not emphasize the statistical significance of our findings. The standard errors are not corrected for multiple hypothesis testing.[39]

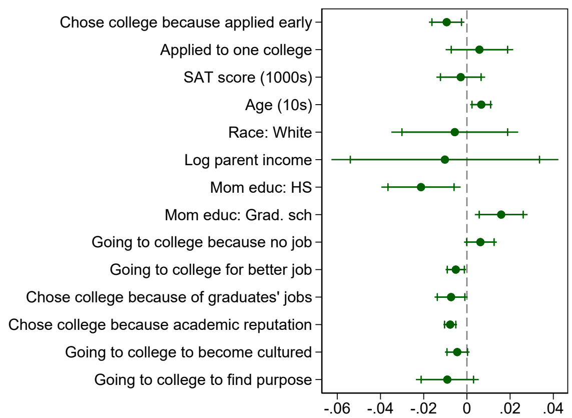
The evidence in Panel (a) is consistent with the regional universities attracting more local students who are at the margin of going to college, whether for financial, social, or academic reasons. This is consistent with our results in Figure 2 that more students in normal school

[36] Appendix Table A46 shows differential likelihood of home-university distance within each of the distance bins listed on the survey. The within-10-mile increase shown in Column (1) of Table 4 corresponds to a decline in going to college between 10 and 100 miles away.

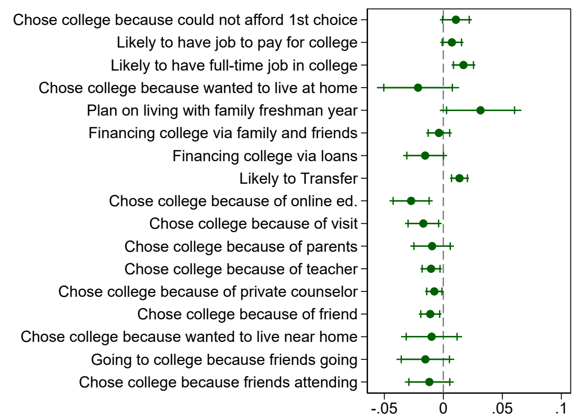
[37] In Appendix E.3, we show the results for an indicator variable just using "Very important."

[38] Following Bond and Lang (2019), we are also clear that we are comparing the fraction of people in normal school and asylum counties who subjectively consider the variable of interest at least somewhat important. There may be differences in the reporting functions across these types of counties that make further generalizations ranking the importance of this variable across groups more difficult.

[39] In fact, in this section, we include only a subset of results that we looked at because we felt they were most elucidating of our mechanisms. All the questions that we looked at can be found in Appendix E.3. We did not include them all in this section because many questions are at least partially redundant and the sheer number of results would make it harder to interpret. We think the qualitative conclusions that we suggest based on the results in the main text are generally consistent with all the results in the appendix.



(a) Characteristics of Students



(b) Geographic Frictions of College Attendance

Figure 5: Differential answers to The Freshman Survey by students who grew up in normal school counties relative to same-state students who grew up in asylum counties. For questions that are answered on a five point scale, we create a dummy variable if the student answered that the reason was "Very important/good" or "Somewhat important/good." Spikes are 95 percent confidence intervals and cross-hatches are 90 percent confidence intervals. Standard errors clustered by state.

counties go to college. In particular, Panel (a) shows that students are less likely to have chosen their college because they applied early ($p < .05$), while at the same time being slightly more likely to have applied to only one college (not significant). Demographically, college freshmen from normal school counties are significantly more likely to be older ($p < .05$). While parental incomes look similar (the point estimate suggests students from normal school counties have slightly lower parental incomes), their parents are slightly more likely to have a graduate degree ($p < .05$), which is one of the few variables suggesting these students are less on the margin of going to college.[40]

Differences in the reasons for going to college are also consistent with students from normal school counties being more marginal college students. Students from normal school counties are less likely to say they chose their college based on the graduates' jobs ($p < .05$), the school's academic reputation ($p < .05$), and they are less likely to say they are going to college to get a better job ($p < .05$), to become cultured or to find purpose (not significant).

[40]Interestingly, splitting the sample by low- and high-parental income does not affect the results in Figure 5 very much (Appendix E.5).

However, they are more likely to be going to college because they do not currently have a job (not significant).

These estimates are additional evidence that the education effects in Section 2.3 are not driven by more economically-mobile families moving into normal school counties prior to college. If that were the case, we might expect students growing up in normal school counties to look like they are less on the margin of going to college.

In Panel (b), we use the survey to potentially disentangle why regional universities are better at reaching students in their own counties than in the same-state asylum counties. We can think of three main hypotheses for why this geographic friction exists. First, having a university nearby may relax the financial burden of attending college, especially if the student can live at home. Second, having a university nearby may decrease information frictions about colleges. Third, people may simply prefer to live near home.

We see that freshmen from normal school counties are more likely to say that they could not afford their first choice (not significant), consistent with the hypothesis that the nearby university eases financial constraints. Consistent with this, we see these students also say that they are more likely to be working during college ($p < .05$). Furthermore, even though students from normal school counties are less likely to say they chose the college because they wanted to live at home (not significant), we see that they are in fact more likely to plan on living at home ($p < .1$), consistent with living at home to reduce costs. Also consistent with sensitivity to cost, students are less likely to be paying for college via family and friends or loans (not significant). We also see that they rate it as more likely that they will end up transferring ($p < .05$). This could be because the nearby regional university is a more affordable way to get the first few years of college education, especially if students live at home, while still preserving the option of a degree from somewhere else. Consistent with this story, we see that students are less likely to say they chose a college because of online education ($p < .05$), suggesting online education may be a substitute for proximity, possibly due to cost-saving. These results relate to a large literature on credit constraints and education, including the Lochner and Monge-Naranjo (2012) review.

We additionally see some evidence that growing up next to a regional university reduces information frictions about college. Freshmen from normal school counties are less likely to say they chose their college because of a visit ($p < .05$), parents (not significant), a teacher ($p < .05$), a private counselor ($p < .05$), or a friend ($p < .05$). These students may not rely on these sources because they already have sufficient information about their local college.

We do not find evidence that location preferences are a major geographic friction. Students are less likely to say that they chose the college because they wanted to live near home (not significant). In addition, we look at whether they are influenced by their friends' choices, which is likely related to their preferences for staying near home. Students from normal school counties are less likely to say that they chose their college, or are going to college at all, because of their friends going (not significant).

4 Conclusion

Regional public universities were established to improve access to higher education in their local communities, thereby improving economic and social mobility. Using a novel strategy and rich data from The Census Tree and Opportunity Insights, we show that regional public universities do have these impacts on their counties, with effects on high school graduation and college attainment for children growing up in the county. We see these effects both for children in 1920, when normal schools had largely become teachers colleges offering a bachelor's in education, and for children in the 1970s and 1980s when the institutions had become comprehensive regional public universities. We further see effects today on employment, household income, marriage, and geographic mobility for children growing up in the county. These effects are large for children from lower-income families. We also show suggestive evidence that these causal effects on the counties are driven by causal effects on people, rather than operating only through sorting.

Our evidence, including from The Freshman Survey, suggests geographic proximity lowers frictions in college attendance, through lowering the cost of college (e.g., by living at home) as well as reducing information frictions. The effects do not appear driven by effects of normal school assignment on the local economy, as there are few differences today in local economic characteristics. Impacts on economic mobility, without impacts on local economic characteristics, are consistent with children attending the local college and then moving away. Indeed we find evidence of increases in geographic mobility from childhood commuting zone.

Our results present important questions for policymakers and future research. The local impact of these universities raises the question of whether they are located optimally, if their objective is to help low-income individuals. We showed that these universities are located in communities with underrepresentation of the lowest-income families, and over-representation of middle-income families. Expanding to lower-income communities will likely have general equilibrium effects, but this seems like an important area for future consideration. Second, how should policymakers address children who do not benefit from proximity to regional universities? Our analysis suggests proximity to a regional university reduces financial costs and information frictions about college, both of which suggest the potential for policy to target assistance to students in underserved areas.

References

Aghion, P., L. Boustan, C. Hoxby, and J. Vandenbussche. 2009. "The Causal Impact of Education on Economic Growth." *Working Paper.*

Alm, James, and John V Winters. 2009. "Distance and intrastate college student migration." *Economics of Education Review*, 28(6): 728–738.

American Council on Education. 1952. *American universities and colleges sixth edition 1952.* Washington, D.C.:American Council on Education.

Anderson, Michael L. 2008. "Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American statistical Association*, 103(484): 1481–1495.

Andersson, Roland, John M. Quigley, and Mats Wilhelmson. 2004. "University decentralization as regional policy: the Swedish experiment." *Journal of Economic Geography*, 4(4): 371–388.

Andrews, Michael. 2021. "How do Institutions of Higher Education Affect Local Invention? Evidence from the Establishment of U.S. Colleges." *American Economic Journal: Economic Policy.*

Andrews, Rodney J, Scott A Imberman, and Michael F Lovenheim. 2020. "Recruiting and supporting low-income, high-achieving students at flagship universities." *Economics of Education Review*, 74: 101923.

Bailey, Martha, Connor Cole, and Catherine Massey. 2020. "Simple strategies for improving inference with linked data: A case study of the 1850–1930 IPUMS linked representative historical samples." *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 53(2): 80–93.

Barro, Robert J, and Xavier Sala-i Martin. 1992. "Convergence." *Journal of political Economy*, 100(2): 223–251.

Bartik, Tim, and George Erickcek. 2008. "The Local Economic Impact of "Eds & Meds": How Policies to Expand Universities and Hospitals Affect Metropolitan Economies." *Metro Economy Series for the Metropolitan Policy Program at Brookings.*

Bedard, Kelly. 2001. "Human capital versus signaling models: university access and high school dropouts." *Journal of Political Economy*, 109(4): 749–775.

Bond, Timothy N, and Kevin Lang. 2019. "The sad truth about happiness scales." *Journal of Political Economy*, 127(4): 1629–1640.

Cantoni, Davide, and Noam Yuchtman. 2014. "Medieval Universities, Legal Institutions, and the Commercial Revolution." *Quarterly Journal of Economics.*

Card, David. 1993. "Using geographic variation in college proximity to estimate the return to schooling."

Card, David, Ciprian Domnisoru, and Lowell Taylor. 2022. "The intergenerational transmission of human capital: Evidence from the golden age of upward mobility." *Journal of Labor Economics*, 40(S1): S39–S95.

Census Office, U.S. Department of the Interior. 1860. "Eighth Census, United States–1860: Instructions to U.S. Marshals."

Centers for Disease Control and Prevention. 1988. "County Cross Reference File." `https://wonder.cdc.gov/wonder/sci_data/codes/fips/type_txt/cntyxref.asp`, Accessed 7/4/23.

Chetty, Raj, and Nathaniel Hendren. 2018. "The impacts of neighborhoods on intergenerational mobility II: County-level estimates." *The Quarterly Journal of Economics*, 133(3): 1163–1228.

Chetty, Raj, John N. Friedman, Emmanuel Saez, Nicholas Turner, and Danny Yagan. 2020. "Data for Income Segregation and Intergenerational Mobility Across Colleges in the United States." *Quarterly Journal of Economics*, 135(3). `https://opportunityinsights.org/paper/undermatching/`.

Chetty, Raj, John N Friedman, Nathaniel Hendren, Maggie R Jones, and Sonya R Porter. 2018. "The opportunity atlas: Mapping the childhood roots of social mobility." National Bureau of Economic Research.

Chetty, Raj, Nathaniel Hendren, Frina Lin, Jeremy Majerovitz, and Benjamin Scuderi. 2016. "Childhood environment and gender gaps in adulthood." *American Economic Review*, 106(5): 282–88.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. 2014. "Where is the land of opportunity? The geography of intergenerational mobility in the United States." *The Quarterly Journal of Economics*, 129(4): 1553–1623.

Coleman, Charles H. 1950. "Eastern Illinois State College-Fifty Years of Public Service."

Collins, William J, and Marianne H Wanamaker. 2022. "African American intergenerational economic mobility since 1880." *American Economic Journal: Applied Economics*, 14(3): 84–117.

Commission to Survey Higher Educational Facilities in Illinois. 1945. "Report of the Commission to Survey Higher Educational Facilities in Illinois."

Council of State Governments. 1950. *The Mental Health Programs of the Forty-Eight States: A Report to the Governor's Conference.* Chicago, IL: Council of State Governments.

Crisp, Gloria, Kevin R McClure, and Cecilia M Orphan. 2021. *Unlocking Opportunity Through Broadly Accessible Institutions.* Routledge.

Derenoncourt, Ellora. 2022. "Can you move to opportunity? Evidence from the Great Migration." *American Economic Review*, 112(2): 369–408.

Do, Chau. 2004. "The effects of local colleges on the quality of college attended." *Economics of Education Review*, 23(3): 249–257.

Douglass, John Aubrey. 2007. *The conditions for admission: Access, equity, and the social contract of public universities.* Stanford University Press.

Doyle, William R, and Benjamin T Skinner. 2016. "Estimating the education-earnings equation using geographic variation." *Economics of Education Review*, 53: 254–267.

Dunham, E Alden. 1969. *Colleges of the Forgotten Americans. A Profile of State Colleges and Regional Universities.* McGraw-Hill Book Company.

Dynarski, Susan, CJ Libassi, Katherine Michelmore, and Stephanie Owen. 2021. "Closing the gap: The effect of reducing complexity and uncertainty in college pricing on the choices of low-income students." *American Economic Review*, 111(6): 1721–56.

Eckert, Fabian, Andrés Gvirtz, Jack Liang, and Michael Peters. 2020. "A Method to Construct Geographical Crosswalks with an Application to US Counties since 1790." *NBER Working Paper*, , (w26770).

Feng, Andy, and Anna Valero. 2020. "Skill-Biased Management: Evidence from Manufacturing Firms." *Economic Journal*, 130(May).

Foster, E.M., H.G. Badger, M.J.S. Carr, B.K. Choate, M. Farr, R.M. Smith, F.J. Kelly, W.J. Greenleaf, and Office of Education (ED) United States Department of the Interior. 1937. *Biennial Survey of Education in the United States,* *1932-1934. Bulletin, 1935, No. 2. Chapter IV: Statistics of Higher Education, 1933-34.* ERIC Clearinghouse.

Fryar, Alisa Hicklin. 2015. "The Comprehensive University: How it Came to Be and What it is Now." In *The University Next Door: What is a Comprehensive University, Who does it Education, and Can it Survive?.* , ed. Mark Schneider and KC Deane. New York, NY:Teachers College Press.

Furbush, E. M., Pollock H. M. (Horatio Milo), A. Veronica Hagan, W. C. (William Chamberlin) Hunt, and United States Bureau of the Census. 1926. *Patients in hospitals for mental disease, 1923.* Washington, D.C.: Government Printing Office. *HathiTrust Digital Library,* https://catalog.hathitrust.org/Record/002085507.

Garin, Andrew, and Jonathan Rothbaum. 2022. "The Long-Run Impacts of Public Industrial Investment on Regional Development and Economic Mobility: Evidence from World War II." *Working Paper.*

Goldsmith-Pinkham, Paul, Peter Hull, and Michal Kolesár. 2024. "Contamination bias in linear regressions." *American Economic Review*, 114(12): 4015–4051.

Grob, Gerald. 2008. *Mental Institutions in America: Social Policy to 1875.* Routledge.

Hausmann, Naomi. 2020. "University Innovation and Local Economic Growth." mimeo.

Higher Education Research Institute. 2023. "CIRP Freshman Survey." Accessed August 2, 2023, https://heri.ucla.edu/cirp-freshman-survey/.

Hoopes, Lauren. 2015. "On the Periphery: A Survey of Nineteenth-Century Asylums in the United States." *All Theses*, 2123.

Howard, Greg, Russell Weinstein, and Yuhao Yang. 2024. "Do universities improve local economic resilience?" *Review of Economics and Statistics*, 106(4): 1129–1145.

Humphreys, Harry Christopher. 1923. *The Factors Operating in the Location of State Normal Schools.* Teachers College, Columbia University.

Indiana Business Research Center. 2020. "StatsAmerica." `http://www.statsamerica.org/CityCountyFinder/`.

Jepsen, Christopher, and Mark Montgomery. 2009. "Miles to go before I learn: The effect of travel distance on the mature person's choice of a community college." *Journal of Urban Economics*, 65(1): 64–73.

Kane, Thomas J, and Cecilia Elena Rouse. 1995. "Labor-Market Returns to Two-and Four-Year College." *The American Economic Review*, 85(3): 600.

Kantor, Shawn, and Alexander Whalley. 2014. "Knowledge Spillovers from Research Universities: Evidence from Endowment Value Shocks." *Review of Economics and Statistics*, 96(1).

Kantor, Shawn, and Alexander Whalley. 2019. "Research Proximity and Productivity: Long-Term Evidence from Agriculture." *Journal of Political Economy*, 127(2).

Kirkbride, Thomas Story. 1854. *On the construction, organization and general arrangements of hospitals for the insane. Sabin Americana, 1500-1926.*, Philadelphia:Lindsay & Blakiston.

Kling, Jeffrey R. 2001. "Interpreting instrumental variables estimates of the returns to schooling." *Journal of Business & Economic Statistics*, 19(3): 358–364.

Klor de Alva, Jorge. 2019. "Is the university next door the way to upward mobility?"

Kozakowski, Whitney. 2023. "Are Four-Year Public Colleges Engines for Economic Mobility? Evidence from Statewide Admissions Thresholds."

Labaree, David F. 2008. "An Uneasy Relationship: The History of Teacher Education in the University." In *Handbook of Research on Teacher Education: Enduring Issues in Changing Contexts (3rd ed.).* , ed. Marilyn Cochran-Smith, Sharon Feiman Nemser and John McIntyre. Washington, DC:Association of Teacher Educators.

Lentz, Eli Gilbert. 1955. *Seventy Five Years in Retrospect: From Normal School to Teachers College to University, Southern Illinois University, 1874-1949.* Vol. 2, University Editorial Board, Southern Illinois University.

Living New Deal. 2024. "Colleges and Universities Sites." `https://livingnewdeal.org/new-deal-categories/education-health/colleges/`, Accessed 12/12/24.

Lochner, Lance, and Alexander Monge-Naranjo. 2012. "Credit constraints in education." *Annu. Rev. Econ.*, 4(1): 225–256.

Long, Bridget Terry. 2004. "How have college decisions changed over time? An application of the conditional logistic choice model." *Journal of econometrics*, 121(1-2): 271–296.

Long, Jason, and Joseph Ferrie. 2013. "Intergenerational occupational mobility in Great Britain and the United States since 1850." *American Economic Review*, 103(4): 1109–1137.

Manson, Steven, Jonathan Schroeder, David Van Riper, Katherine Knowles, Tracy Kugler, Finn Roberts, and Steven Ruggles. 2023. "IPUMS National Historical Geographic Information System: Version 18.0 [dataset]." http://doi.org/10.18128/D050.V18.0.

Maxim, Robert, and Mark Muro. 2020. "Restoring Regional Public Universities for Recovery in the Great Lakes." *Brookings Metropolitan Policy Program.*

Mayhew, Lewis B. 1969. "Long Range Planning for Higher Education. Studies in the Future of Higher Education."

McClure, Kevin R., and Alisa Hicklin Fryar. 2020. "Vulnerable, Fragile, and Endangered? Financial Rhetoric and Reality at Regional Public Universities." *New Directions for Higher Education*, 190, Summer.

Moretti, Enrico. 2004. "Estimating the Social Return to Higher Education: Evidence from Longitudinal and Repeated Cross-Sectional Data." *Journal of Econometrics*, 121.

Mountjoy, Jack. 2022. "Community colleges and upward mobility." *American Economic Review*, 112(8): 2580–2630.

National Association of Counties. 2023. "County Explorer." `https://explorer.naco.org/?find=true`, Accessed 12/18/23.

Ogren, Christine A. 2003. "Rethinking the "nontraditional" student from a historical perspective: State normal schools in the late nineteenth and early twentieth centuries." *The Journal of Higher Education*, 74(6): 640–664.

Ogren, Christine A. 2005. *The American State Normal School: An Instrument of Great Good.* Springer.

Price, Joseph, Kasey Buckles, Adrian Haws, and Haley Wilbert. 2023*a*. "The Census Tree, 1850-1860." https://doi.org/10.3886/E193225V1.

Price, Joseph, Kasey Buckles, Adrian Haws, and Haley Wilbert. 2023*b*. "The Census Tree, 1920-1940." https://doi.org/10.3886/E193238V1.

Pryor, John H., Sylvia Hurtado, Linda DeAngelo, Laura Palucki Blake, and Serge Tran. 2010. "The American Freshman: National Norms Fall 2010."

Quinton, Sophie. 2020. "COVID-19 Could Be End of Line For Some Regional Colleges." *Stateline, The Pew Charitable Trusts*. https://www.pewtrusts.org/en/research-and-analysis/blogs/stateline/2020/06/04/covid-19-could-be-end-of-line-for-some-regional-colleges, Accessed 10/25/20.

Romano, Joseph P, and Michael Wolf. 2005. "Stepwise multiple testing as formalized data snooping." *Econometrica*, 73(4): 1237–1282.

Ruggles, Steven, Catherine A. Fitch, Ronald Goeken, J. David Hacker, Matt A. Nelson, Evan Roberts, Megan Schouweiler, , and Matthew Sobek. 2021*a*. "IPUMS Ancestry Full Count Data: Version 3.0 [dataset]." https://doi.org/10.18128/D014.V3.0.

Ruggles, Steven, Sarah Flood, Sophia Foster, Ronald Goeken, Jose Pacas, Megan Schouweiler, and Matthew Sobek. 2021*b*. "IPUMS USA: Version 11.0 [dataset]." https://doi.org/10.18128/D010.V11.0.

Russell, Lauren C., and Michael J. Andrews. 2022. "Not the Great Equalizer? Local Economic Mobility and Inequality Effects from the Establishment of U.S. Universities."

Russell, Lauren, Lei Yu, and Michael Andrews. 2022. "Higher Education and Local Educational Attainment: Evidence from the Establishment of US Colleges." *Review of Economics and Statistics*.

Schneider, Mark, and KC Deane. 2015. "Introduction." In *The University Next Door: What is a Comprehensive University, Who does it Education, and Can it Survive?*. , ed. Mark Schneider and KC Deane. New York, NY:Teachers College Press.

Seltzer, Rick. 2019. "Squeezed From All Sides: Opportunities and Challenges for Regional Public Universities." *Inside Higher Ed*.

Seltzer, Rick. 2020. "Pa. State System Moves Forward with Modified Merger Plan." *Inside Higher Ed*.

Smith, Jonathan, Joshua Goodman, and Michael Hurwitz. 2020. "The economic impact of access to public four-year colleges." National Bureau of Economic Research.

Soltow, Lee, and Edward Stevens. 1981. *The Rise of Literacy and the Common School in the United States: A Socioeconomic Analysis to 1870.* University of Chicago Press.

StatsAmerica. 2023. "City-to-County Finder." `https://www.statsamerica.org/CityCountyFinder/`, Accessed 12/18/23.

Tan, Hui Ren. 2023. "A different land of opportunity: The geography of intergenerational mobility in the early twentieth-century united states." *Journal of Labor Economics*, 41(1): 77–102.

United States Department of Education. National Center for Education Statistics. 1998. *Higher Education General Information Survey (HEGIS), 1970: Fall Enrollment.*

United States Department of Education. National Center for Education Statistics. 1999. *Higher Education General Information Survey (HEGIS), 1975: Fall Enrollment.*

U.S. Bureau of Economic Analysis. February 2025. "Regional Economic Accounts." Retrived from: `https://apps.bea.gov/regional/downloadzip.htm`.

U.S. Census Bureau. 2002. "Federal Information Processing System (FIPS) Codes for States and Counties." `https://www2.census.gov/programs-surveys/popest/technical-documentation/methodology/1990-2000/90s-fips.txt`, Accessed 12/18/23.

U.S. Department of Education, National Center for Education Statistics. 2020. "Integrated Postsecondary Education System (IPEDS)." `https://nces.ed.gov/ipeds/`.

U.S. Department of Education, National Center for Education Statistics. 2021. "Baccalaureate and Beyond: 2008/2018 (B&B)." Computation by NCES PowerStats on 8/19/22.

Valero, Anna, and John Van Reenen. 2019. "The Economic Impact of Universities: Evidence from Across the Globe." *Economics of Education Review*, 68.

Wendler, Walter V. 2018. "Regional University Focus." *Public Purpose*, Winter.

Willingham, Warren W. 1970. "Free-Access Higher Education."

Young, Alwyn. 2020. "RANDCMD: Stata module to compute randomization inference p-values."

Zimmerman, Seth D. 2014. "The returns to college admission for academically marginal students." *Journal of Labor Economics*, 32(4): 711–754.

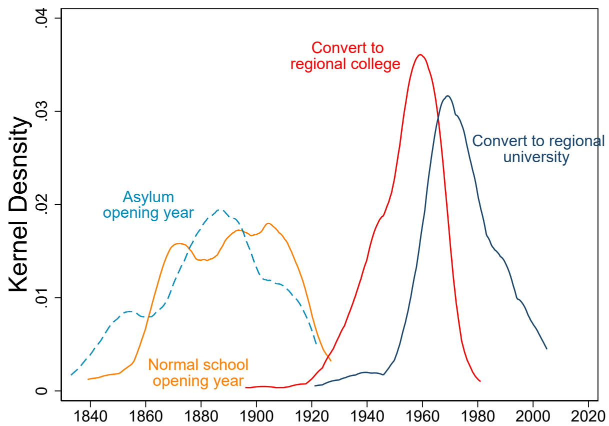
A Normal School and Asylum County History

In this appendix, we present further details on the assignment of normal schools and asylums, balance on 1850 county characteristics, and the impact on the higher-education sector.

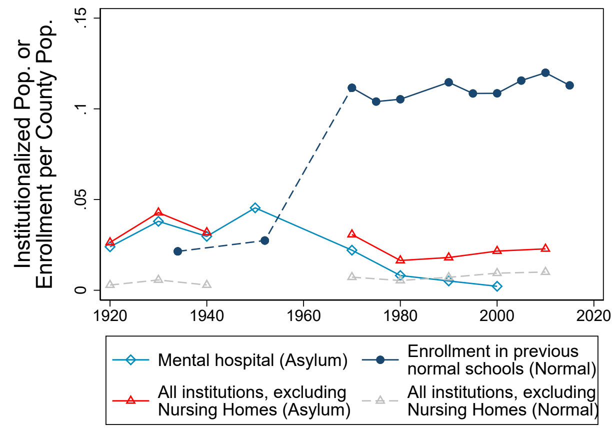
In Figure A1, which is reproduced from Howard, Weinstein and Yang (2024), we show the timeline of the opening and conversion of normal schools, compared to asylum counties (Panel a), as well as the statistics on the size of these institutions over time (Panel b).[41] We also include a map of the institutions to show that both normal schools and asylums were common across the entire country (Panel c). As we document in Howard, Weinstein and Yang (2024), university-level enrollment data in the 1933-1934 academic year comes from the *Biennial Survey of Education, 1932-1934* (Foster et al., 1937), and in 1952 from *American Universities and Colleges, Sixth Edition* (American Council on Education, 1952). Total university enrollments in 1970 and 1975 are from the *Higher Education General Information Survey (HEGIS)* (United States Department of Education. National Center for Education Statistics., 1998, 1999), and from 1980 to 2015 are from IPEDS (U.S. Department of Education, National Center for Education Statistics, 2020). Institutional population by institution type are from the decennial censuses of 1920 through 1940 using 100% counts from IPUMS (Ruggles et al., 2021*b*). Because 100% counts are not available in 1950, we instead collected asylum-level resident population data from the Council of State Governments (1950).

Table A1 shows balance on characteristics in 1850. Normal school counties are smaller in population in 1850, and there is some evidence they are less urban, have a greater fraction of farmers, and have lower real estate values per capita in 1850. We note that while there is not a statistically significant difference in 1850 population levels, using log population yields a coefficient of -.30, statistically significant at the 10% level. As we show in Howard, Weinstein and Yang (2024), there is not a statistically significant difference between normal school and asylum counties in log population in 1920, when we have data on all states. There is also not a significant difference in log population in 1840. Finally, we note that there is an extreme outlier in terms of 1850 population: New York County. Omitting this county yields a statistically insignificant coefficient of -881 in row 1, and a mean of 26,167 in column 2 row 1. The last row shows there is no statistically significant difference in the first principal component of the variables in Table A1, in which all variables have positive loadings except percent non-white slave population and percent farmer.

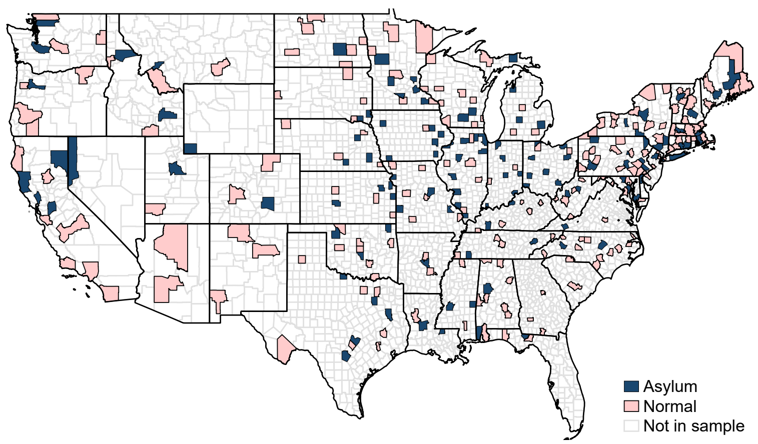
[41] The figures and tables we present in this paper based on Howard, Weinstein and Yang (2024) are not perfectly identical to those in Howard, Weinstein and Yang (2024), because of a corrected classification of two asylum counties from that paper. The differences are very slight. County FIPS 29043 had been classified as an asylum county when it should have been 21047, and 36117 had been classified as an asylum county when it should have been 37191. However, 37191 is also normal school county. Thus, our sample includes 204 normal school counties and 125 asylum counties.



(a) Asylum and Normal School Opening Years



(b) Asylum and Normal School Size Over Time



(c) Locations of Normal Schools and Asylums

Figure A1: **History of Normal Schools and Insane Asylums.** *Notes*: Figure (a) shows opening years for normal schools and asylums. We use an Epanechnikov kernel with a five-year bandwidth for density estimation. The year in which previous normal schools convert to state colleges and state universities is defined to be the year that the school's name changes to college and university respectively. Figure (b) shows average enrollment in normal schools (or in colleges that had been normal schools) per county population in normal school counties. We also show average institutionalized population per county population for both normal school and asylum counties. Depending on the year, institutionalized population includes population in mental institutions, correctional institutions, institutions for the elderly, handicapped, and poor, juvenile facilities, and nursing/skilled nursing facilities. College enrollment in Maine and Vermont is missing in 1952; however, using a balanced sample yields a similar figure. Figure (c) shows a map of the locations of the normal school and asylum counties in our sample. See text for data sources.

Table A1: County Characteristics in 1850

	Normal	Asylum	Within-State Difference
Population, 1850	27,316	31,730	-7,016
	[33,125]	[58,955]	(7,094)
Proportion of population, 1850:			
Urban, Places 2500 and over	0.11	0.15	-0.05^{+}
	[.2]	[.23]	(0.03)
In cities, 25,000 and over	0.03	0.05	-0.02
	[.13]	[.19]	(0.02)
Non-white, free	0.02	0.02	-0.00
	[.03]	[.03]	(0.00)
Non-white, slave	0.14	0.09	0.00
	[.2]	[.17]	(0.01)
Farmer	0.25	0.25	0.03^{+}
	[.12]	[.12]	(0.01)
Real estate value per capita	239.21	252.8	-19.42^{+}
	[136.47]	[188.5]	(11.14)
First principal component			-0.37
			(0.24)

Notes: Columns 1 and 2 show mean and standard deviation of county characteristics for normal school and asylum counties. Column 3 shows the coefficient on normal school county, when the dependent variable is the county characteristic, and we include state fixed effects. We show standard errors clustered at the state level in parentheses in column 3. There are 139 normal school counties and 87 asylum counties. We restrict the 1850 samples to counties in states that had entered the Union by the day of the census in 1850. We use the Eckert et al. (2020) crosswalk to 1990 counties. When using log population in 1850 as the dependent variable in column 3, the coefficient on normal school county is -.30, statistically significant at the 10% level. Fraction of the population that is a farmer is the fraction of the population who are at least 15, and not living in group quarters. Real estate value per capita is the sum of all real estate value owned by individuals in the county (not living in group quarters), divided by the total non-group-quarters population. The last row shows a regression in which the dependent variable is the first principal component of the above variables, in which all variables have positive loadings except percent non-white slave population and percent farmer. See Howard, Weinstein and Yang (2024) for balance on other variables in 1840 and in 1920.
$^{+}$ $p < 0.1$, * $p < .05$, ** $p < .01$

A.1 Effect of Normal School Assignment on the Local Higher Education Sector

Table A2, reproduced from Howard, Weinstein and Yang (2024), shows the effects of normal schools on the local higher education sector, showing that normal school counties have more public four-year colleges, and the colleges have higher enrollment and more degrees awarded per population. The normal school counties also have a higher share of the population with a bachelor's degree. While there are some insignificant negative effects on other types of universities, these universities are small, so the net effect is still a much larger university presence, when measured by enrollment or degrees, even if that is not the case when measured by the total number of colleges.

Table A2: County-level Higher Education Sector, 1980

	(1)	(2)	(3)
	Variable Means		Difference in Means With State FE
	Normal	Asylum	(1) - (2)
Has regional college formerly normal school	0.91	0.00	0.93**
	(0.28)	(00)	(0.02)
Total public four-year colleges	1.11	0.45	0.69**
	(0.67)	(0.88)	(0.13)
Total private four-year colleges	1.39	1.96	-0.48
	(3.27)	(4.63)	(0.55)
Total two-year colleges	0.97	1.17	-0.24
	(2.17)	(2.17)	(0.31)
Enrollment as % of population	11.72	4.60	8.48**
	(9.23)	(5.52)	(1.61)
Full-time enrollment as % population	8.52	2.99	6.51**
	(7.4)	(4.35)	(1.27)
Total degrees awarded as % of population	3.04	0.94	2.48**
	(2.77)	(1.41)	(0.5)
Bachelor's degrees awarded as % of population	1.43	0.39	1.23**
	(1.38)	(0.69)	(0.25)
% Population over 25 with Bachelor's degree	16.57	15.05	2.04*
	(4.79)	(6.11)	(0.86)
% Population over 25 with 1-3 years college	15.40	15.03	0.55
	(3.89)	(3.98)	(0.35)

Notes: Source: Howard, Weinstein and Yang (2024). Columns (1) and (2) show means and standard deviations in parentheses. For panel A, column (1) includes 204 normal school counties, and column (2) includes 125 asylum counties. Panel A data are constructed using IPEDS, except the bachelor's share and some-college share which are from the census, obtained from NHGIS. Column (3) displays coefficients from regressing each variable on the normal school county indicator with state fixed effects, clustering standard errors at the state level. $^{+}$ $p < 0.1$, * $p < .05$, ** $p < .01$.

Table A3 further highlights differences between universities in normal school and asylum counties, in terms of cost, completion rates, and selectivity.

We use data from the IPEDS complete data files in 2009. Of the 329 normal school and asylum counties in our data, 51 do not have four-year colleges that are in our roster of Title IV colleges and universities from IPEDS in 2009. Of those 51, 41 are asylum counties. The analysis below shows differences in university characteristics between the universities in normal school counties and asylum counties. In addition to these differences, this difference in the extensive margin is also important for interpreting the results in our paper.

The evidence we present below in Table A3 is consistent with regional public universities raising access among local students by having lower financial costs and lower admissions selectivity. We also see that universities in normal school counties have lower completion rates, which is especially important given our finding in Figure 2 that children growing up in normal school counties are still more likely to get a degree. We describe the data and analysis in detail below.

For the cost variables, we use the financial aid and net price surveys in 2009, which was the first year that variables related to net price, as well as the variables denoting aid by income levels, were available. These variables are reported for the 2008-2009 academic year. We obtain the tuition and fee data from the institutional characteristics survey, which reports these variables for the 2009-2010 academic year. We restrict to four-year colleges and universities in order to understand the differences in local options for students who are considering a four-year degree.

We focus on the following which are reported for full-time first-time degree or certificate seeking students (and for public universities additionally restricted to students paying the in-district or in-state tuition rate): average amount of grant and scholarship aid in all income levels for students receiving Title IV federal student aid, average amount of grant and scholarship aid in income level (0-30,000) for students receiving Title IV federal student aid, average amount of grant and scholarship aid in income level (30,001- 48,000) for students receiving Title IV federal student aid, average net price for students receiving grant or scholarship aid, average net price for students in income level (0-30,000) receiving Title IV federal financial aid, and average net price for students in income level (30,001-48,000) receiving Title IV federal financial aid. Average net price is equal to the sum of published tuition and required fees, books and supplies, and the weighted average room and board and other expenses minus average amount of aid. Details can be found in the IPEDS documentation for the Student Financial Aid Data File, which can be accessed through the Complete Data Files section of the IPEDS website. We also analyze the sum of in-state average tuition for full-time undergraduates and in-state required fees for full-time undergraduates.

For the analysis of these cost variables, we restrict to institutions for which the variables above are all non-missing, along with the number of full-time first-time degree/certificate seeking undergraduates in the financial aid cohort. This facilitates interpreting differences between tuition and aid, which otherwise could be due to differences in the composition of universities reporting these variables. In practice this is not very restrictive as most universities report all of the variables above. There are many more universities that do not report the variables for students with incomes higher than $48,000, and so we do not show those results.

We construct county-level averages of these variables weighting the institutions in the county by the number of full-time first-time degree/certificate seeking undergraduates in the financial aid cohort.

Universities in normal school counties are 44 percentage points more likely to be public (Table A3). Tuition and fees are roughly 50% lower than tuition and fees at universities in asylum counties within the same state, but the average aid is also roughly 50% lower. Average net price is about 25% lower for students at universities in normal school counties who are receiving aid, and 18-22% lower for poor students receiving Title IV federal financial aid. For students with income level (30,001-48,000) this is a difference of approximately 2500 dollars per year, given the mean net price for students at this income level in asylum counties of roughly 14,200 dollars.

For bachelor's degree completion, we use the number of bachelor's degree-seeking students who were enrolled as full-time, first-time students in 2003 and completed a bachelor's degree within 150% of normal time, as a fraction of the adjusted starting cohort of these students. Details are available in the IPEDS file documentation for graduation rate data, which can be downloaded from the IPEDS Complete Data Files website. We construct county-level averages, weighting observations by the adjusted starting cohort of these students, and include only the four-year institutions. The average bachelor's degree completion rate within 150% of normal time is 4.2 percentage points lower for universities in normal school counties, statistically significant at the 5% level (Table A3). This is roughly 8% lower given the mean completion rate of 56% in asylum counties.

For selectivity variables, we use the 2009 Admissions and Test Scores survey from IPEDS and focus on several variables. First, we use whether the institution reports having an open admission policy for all or most of its undergraduate students. Institutions that do not have open admissions are then asked for the number of applications and number of admitted students, as well as how letters of recommendation and test scores are used in the admission process. They can respond that they are required, recommended, or neither required nor recommended. We create an indicator for whether these are either required or recommended. Institutions that do not have open admissions and that require test scores, and for which at least 60% of the entering students reported scores for the test, are then asked for the 25th and 75th percentiles of the SAT and ACT test score distributions. We weight observations by the number of first-time enrolled undergraduate degree- or certificate-seeking undergraduate students, and construct the county-level average. Given that these variables are reported by different number of universities based on their admissions requirements, the number of counties in the regressions varies with the dependent variable. For the open admissions regression, we have 271 of our counties, for the fraction admitted and test score requirements we have 252 counties, for the ACT distribution we have 232 counties, for the SAT Verbal distribution we have 211 and for SAT Math we have 213.

Universities in normal school counties are less likely to have open admissions policies, but this is not statistically significant (Table A3). Conditional on having some selectivity in admissions, universities in normal school counties are less likely to require or recommend letters of recommendation, but more likely to require or recommend test scores. Conditional on requiring test scores, and at least 60% of entering students reporting the test, the test scores are lower at universities in normal school counties.

Table A3: Characteristics of Four-Year Universities in Normal School and Asylum Counties

	Normal	Asylum	Diff Normal-Asylum
Public university	0.82	0.39	0.447**
	[.27]	[.38]	(0.051)
Tuition and fees	9767	17350	-0.532**
	[6147]	[9241]	(0.059)
Grant and scholarship aid	6227	9984	-0.485**
	[3867]	[5573]	(0.068)
Net price, students receiving grant or scholarship aid	12748	17021	-0.246**
	[3977]	[4794]	(0.042)
Grant and scholarship aid in income level (0-30,000)	10102	14790	-0.331**
	[5221]	[7677]	(0.057)
Grant and scholarship aid in income level (30,001-48,000)	8377	13110	-0.428**
	[5066]	[7497]	(0.064)
Net price in income level (0-30,000), students receiving Title IV federal financial aid	9732	12535	-0.217**
	[3698]	[3948]	(0.056)
Net price in income level (30,001-48,000), students receiving Title IV federal financial aid	11456	14214	-0.175**
	[3703]	[3787]	(0.043)
BA completion within 150% normal time	0.5	0.56	-0.042*
	[.15]	[.18]	(0.020)
Open admission policy	0.09	0.13	-0.051
	[.27]	[.26]	(0.035)
Conditional on non-open admission			
Fraction admitted	0.68	0.68	-0.002
	[.15]	[.14]	(0.019)
Letters of recommendations required or recommended	0.46	0.67	-0.224**
	[.46]	[.4]	(0.065)
Test scores required or recommended	0.97	0.87	0.111**
	[.07]	[.27]	(0.031)
Conditional on non-open admission, test scores required, and $\geq 60\%$ of enrolled students submitted			
ACT Composite, 25th percentile	19.55	21.2	-1.613**
	[2.47]	[2.6]	(0.273)
ACT Composite, 75th percentile	24.41	26.31	-1.800**
	[2.61]	[2.4]	(0.268)
SAT Verbal, 25th percentile	462.07	481.96	-19.359*
	[50.36]	[53.58]	(7.336)
SAT Verbal, 75th percentile	568.37	598.41	-25.325**
	[51.69]	[49.68]	(7.191)
SAT Math, 25th percentile	467.07	495.04	-26.051**
	[55.25]	[60.46]	(8.152)
SAT Math, 75th percentile	576.77	608.28	-23.838**
	[55.98]	[57.37]	(8.454)

Notes: $^{+}$ $p < 0.1$, * $p < .05$, ** $p < .01$. Column 1 gives the enrollment-weighted average of the variable for four-year universities in normal school counties, and column 2 gives the average in asylum counties. Standard deviations are in brackets in columns 1 and 2. Column 3 presents the coefficient on normal school county when regressing the variable on normal school county and state fixed effects. For tuition, aid, and net price variables, column 3 shows the difference in logs. Standard errors in column 3 are clustered at the state level, and shown in parentheses. See text for details.

A.2 State Income Per Capita in 1929 and Number of Institutions

In this subsection, we analyze whether states with higher per capita incomes in the early 1900s built more normal schools and asylums, and whether states that had been poorer in 1929 had a similar number of regional universities by 1987 compared to wealthier states in 1929, as income per capita across states converged. We tested this using state income per capita data in 1929 from the Bureau of Economic Analysis (U.S. Bureau of Economic Analysis, February 2025). Similar to Barro and Sala-i Martin (1992), we exclude personal current transfer receipts from total personal income. Among states that have both normal schools and asylums, we see that states with higher per capita income in 1929 had built more normal schools and asylums (using the normal schools and asylums in our sample) ($p < .05$) (Table A4). There is also suggestive evidence that they had built more normal schools, with the coefficient slightly less than half of the coefficient when the outcome is normal schools plus asylums. However, this is not significant at the 10% level. Interestingly, the magnitude suggests an even more positive relationship between 1929 state income per capita and total public, non-research universities in the state in 1987 (including the former normal schools), suggesting that the relationship did not weaken with regional convergence over the 20th century.[42] However, we also cannot rule out a coefficient of zero or negative coefficients, so we do not want to overemphasize these results.

Table A4: Institutions and State Income Per Capita, 1929

	(1)	(2)	(3)
	Normal Schools and Asylums	Normal Schools	Public Non-research Univ., 1987
State income per capita, 1929	0.009*	0.004	0.007
	(0.004)	(0.003)	(0.006)
Observations	40	40	40
R-squared	0.176	0.064	0.054
Dependent variable mean	7.850	4.825	9.500
Standard deviation, income per capita	213.410	213.410	213.410

Notes: $^{+}$ $p < 0.1$, * $p < .05$, ** $p < .01$. The dependent variable is the number of institutions in the state by type. There is one observation per state. Robust standard errors are in parentheses.

[42] We use the 1987 Carnegie rating, and proxy for regional universities by summing the total of Doctorate-Granting I, Doctorate-Granting II, Comprehensive I, Comprehensive II, Liberal Arts I, and Liberal Arts II. We do not include Research I or Research II, or specialized universities.

B Historical Differences in Social Mobility

B.1 Differences in Economic Mobility in 1850-1860: Evidence from the Census Tree

In Section 1.3, we argued that there were no pre-existing differences in economic mobility, supporting our identification assumption. In this section, we go into the details of our analysis and present additional results.

We identify children in the 1850 full count census who are living in normal school or asylum counties, not living in group quarters, and living with at least one of their parents based on the IPUMS imputation procedure (Ruggles et al., 2021*a*).[43] We then link these records to their record in the 1860 full count census using the 1850 to 1860 Census Tree crosswalk (Price et al., 2023*a*) and the data from the 1860 full count of the census (Ruggles et al., 2021*a*).[44]

Seven states had already opened 10 normal schools before 1860. We exclude these states from this analysis, as the individuals in 1860 may be affected by the normal schools that had already been opened in their states.[45] We also exclude states that had not yet entered the Union by the day of the 1850 census, and we drop states that did not eventually have at least one normal school and one asylum county.

We use the Eckert et al. (2020) crosswalk from 1850 to 1990 counties to identify individuals living in what are today's normal school and asylum counties. We construct county-level averages for many outcomes pooling men and women, but separating by race.

To study economic mobility, we focus on 16-18 year-olds in 1850. This age group increases the likelihood that we measure real differences in mobility rather than slight differences in life cycles. We discuss this further below. We analyze two types of outcomes related to social mobility. First, we test for differences in adult outcomes that are conditional on low parental socioeconomic status. In addition, we test for differences in occupational mobility, which measures the share of people in a different occupation than their parents, a measure that has been used in the literature including by Long and Ferrie (2013).

We identify children from lower socioeconomic status families in 1850 using the value of real estate owned by their mother and father, based on the IPUMS imputation of family relationships. We focus on children whose parents are approximately in the bottom-third (less than or equal to 150 dollars of real estate value) of the distribution of parents of 16-18 year-old White children in our sample of states described above.[46]

Real estate wealth arguably differentially captures wealth of farmers, relative to people living in more urban areas who may have wage income but not real estate. In the sample

[43] The 1850-1870 censuses did not explicitly ask about family interrelationships.

[44] We downloaded IPUMS data for 6-19 year-olds in 1850, and 13 to 32 year-olds in 1860 in case of inconsistencies in reporting age; however our focus will be on 16-18 year-olds in 1850 for the reasons discussed below.

[45] Ogren (2003) suggests many of the enrolled normal school students were in their 20s.

[46] Real estate wealth of \$150 is the 37th percentile of the parental real estate wealth distribution in the states in our sample, while wealth of zero is the 33rd percentile. We use \$150 to increase sample sizes for these low socioeconomic status children. An alternative would be to use parental literacy as a proxy for parental educational attainment (Soltow and Stevens, 1981). However, only 11 percent of white individuals over age 25 were illiterate in 1850, making our sample sizes too small to carry out the analysis.

described for this section, we see that the fraction of children whose parents are farmers is higher in normal school counties. This could yield misleading conclusions, as children in asylum counties may have similar socioeconomic status based on other non-observed wealth measures that are not skewed towards farmers. Using a relatively low parental wealth threshold avoids some of these issues as the parents have little they could pass on to the children. We do not use an even lower threshold (such as zero real estate value) as this reduces already small sample sizes of 16-18 year-olds in each county.

Finally, there are several normal school and asylum counties with very small sample sizes of 16-18 year-olds in 1850 whose parents' value of real estate is less than or equal to 150 dollars. We drop states in which 50% or more of the normal school or asylum counties have sample sizes less than or equal to 10. This includes four states (Florida, Iowa, Texas, and Wisconsin), all of which entered the Union between 1845 and 1850.[47]

Our sample in this section includes 15 of the 40 states that opened a normal school and asylum at some point during our sample period: Alabama, Arkansas, Indiana, Kentucky, Louisiana, Maine, Maryland, Mississippi, Missouri, New Hampshire, North Carolina, Ohio, Tennessee, Vermont, and Virginia. We show results only for White individuals, as the analysis using county-level averages for free Black inhabitants, and using the same sample restrictions described above, does not yield any states.

We focus on 16-18 year-olds in 1850. We acknowledge that our focus on roughly 26-28 year-olds in 1860 (16-18 in 1850) may not be ideal since these are still relatively young ages for understanding economic mobility. However, linking the children in 1850 to their 1870 census records presents several challenges. First, this census comes right after the Civil War, and the immediate impacts of the war may make it difficult to identify general differences in mobility in the pre-normal school era. Second, an additional nine states open their first normal schools between 1860-1869. This would yield even fewer states on which we can focus our analysis, that had not yet opened their first normal school at the time we measure outcomes. Including individuals older than 18 in 1850 would allow us to observe outcomes for older individuals in 1860, but fewer of these individuals will still be living with their parents in 1850. Our focus on coresident children 18 and younger is similar to Card, Domnisoru and Taylor (2022).

Nationally, the fraction of White males living with at least one of their parents in 1850 is 90% or above through 13-year-olds, among those not living in group quarters. It falls after that, but is still 80% at 17, and 74% at 18. For White women, the rate is also roughly 90% or above for those 13 and under. However, it falls more quickly after that than for males. The rate is 74% for 17 year-olds and 64% for 18 year-olds. In our sample of normal school and asylum counties in the states described above, there are no differences in the fraction of 6-8 year-old White boys or girls in 1850 that are still living with their parents in 1860 in normal school versus same-state asylum counties.

Of 61,600 White males ages 16-18 in the 1850 census who grew up in normal school or asylum counties in the sample of states described above, and living with at least one of their parents, 35,827 (roughly 58%) have links between their 1850 and 1860 census records in the Census Tree.[48] Of 59,129 White females ages 16-18 in the 1850 census who grew

[47] In addition to small sample sizes, the 1850-1990 county crosswalk for these states (except Iowa) shows the county boundaries in 1850 were quite different for some of today's normal school and asylum counties.

[48] We note that some of these individuals do not have links to their 1860 record because they are no longer

up in normal school or asylum counties in the sample of states described above, and living with at least one of their parents, 24,836 (roughly 42%) have links between their 1850 and 1860 census records.[49] We do not see statistically significant differences between normal school and asylum counties in the fraction whose ID is linked to the 1850-1860 Census Tree crosswalk based on their 1850 ID.[50]

We calculate county-level averages for individuals who were children 16-18 years old living with their parents in the county in 1850, and estimate the following county-level regression:

$$y_{c1860} = \beta \text{Normal}_{c1850} + \alpha_{s1850} + \epsilon_c \tag{A1}$$

where Normal_{c1850} is an indicator variable for living in 1850 in a county that will receive a normal school, and y_{c1860} is the outcome of interest in 1860, for the people who grew up in county c in 1850, regardless of their county in 1860. We include state fixed effects, so we compare 1860 outcomes of people who grew up in normal school or asylum counties but whose 1850 state was the same. Given there are 15 states in these regressions, we present unclustered, heteroskedastic-robust standard errors, as well as p-values based on randomization inference (we use the Stata command "permute"), as described in Section 1.3.1.

Table A5 shows differences in occupations of parents in 1850 and children in 1860 in normal school and asylum counties. We code parents as having a given occupation if either the mother or father based on IPUMS imputation procedures have the occupation. Similarly, we code the household occupations of the children in 1860 if either the individual or their spouse have the occupation. This allows us to include both women and men. Table A6 shows the results when focusing only on fathers and sons. There are no statistically significant differences in the occupations of parents or children, for children growing up in normal school versus asylum counties. However, magnitudes suggest children in normal school counties are more likely to have parents who were farmers, and less likely to have parents who were in craft/operative or white-collar occupations. In 1860, magnitudes suggest the children were more likely farmers. The fraction of children with different household occupations than their parents, excluding children for which the household occupation is a non-occupation response, is 1.3 percentage points smaller in normal school counties, though not statistically significant.

Table A6 shows a similar table, but only for fathers and sons. The results are broadly similar, though the magnitudes are larger. In normal school counties, the lower fraction with

living in 1860.

[49] When including only individuals with parents' real estate wealth less than or equal to $150, the fraction merging for men and women is 52% and 34% respectively.

[50] A small number of those who merge to the 1850 ID in the 1850-1860 crosswalk do not merge to an 1860 census record, presumably because we restricted the 1860 full count census file to those who were 13 to 32 in 1860 (to align with the ages we selected for the 1850 census). The link rates above only look at likelihood of merging to the 1850 ID in the crosswalk, which captures the vast majority of the lack of merge to the crosswalk. There were 60,663 White, 16-18 year-olds whose 1850 ID merged to the 1850-1860 Census Tree crosswalk, and 748 of those (1.2%) do not have 1860 census records, presumably because of the age restriction we imposed on the 1860 file. There are also not statistically significant differences in the fraction merging between normal school and asylum counties when including only individuals whose parental wealth is less than or equal to $150.

Table A5: Parent and Child Occupations, Normal School and Asylum Counties

	(1)	(2)	(3)	(4)	(5)	(6)
	Unskilled	Craft/Oper	White Collar	Non Occ	Farmer	Diff from Parent
Panel A: Parent Occupations 1850, Parents of 1850 16-18 Year-Old White Children						
Grew up in Normal School County	-0.008	-0.019	-0.015	0.003	0.039	
	(0.006)	(0.013)	(0.014)	(0.012)	(0.034)	
Observations	102	102	102	102	102	
R-Squared	0.411	0.381	0.138	0.426	0.284	
Mean DV, Asylum Counties	0.041	0.124	0.067	0.143	0.625	
p-value randomization inference	0.259	0.14	0.14	0.816	0.18	
Panel B: Child Occupations 1860, 1850 16-18 Year-Old White Children						
Grew up in Normal School County	0.012	-0.010	-0.009	-0.010	0.027	-0.013
	(0.012)	(0.011)	(0.012)	(0.012)	(0.021)	(0.014)
Observations	102	102	102	102	102	102
R-Squared	0.261	0.563	0.312	0.272	0.541	0.429
Mean DV, Asylum Counties	0.142	0.154	0.093	0.282	0.345	0.563
p-value randomization inference	0.424	0.418	0.333	0.489	0.168	0.466

Notes: $^{+}$ $p < 0.1$, * $p < .05$, ** $p < .01$ Outcomes are county-level averages for individuals who were 16-18 years old and living with at least one parent in 1850, among individuals who could be matched to their 1860 records using The Census Tree (Price et al., 2023*a*). Panel A shows the occupations of these childrens' parents in 1850. We code parents as having an occupation if either the mother or the father has the occupation, using the IPUMS imputed family relationships. Panel B shows the occupations of the children in 1860, when they are roughly 26-28 years old. We show the household occupation, which is equal to one if either the individual or the spouse has the occupation, using the imputed family relationships. Non-occupational response is coded as one only if both parents (panel A) or the individual and their spouse if they have one (panel B) have non-occupational responses. We follow Long and Ferrie (2013) in defining occupational groups: unskilled are service workers and laborers, including farm laborers, white-collar are professional, technical, and kindred; managers, officials, and proprietors; clerical; and sales. Farmers are farm owners and farm managers. In column 6, panel B, we show the fraction of individuals with different household occupation in 1860 than their parents in 1850. For this measure, we exclude the individuals whose household occupation in 1860 was a non-occupational response. We use the 1850-1990 county crosswalk from Eckert et al. (2020). There are 15 states in the regression. Robust standard errors are in parentheses. P-values based on randomization inference are obtained by permuting within states the assignment of normal school counties, among our sample of normal school and asylum counties. This is based on 1000 permutations. See text for details.

parents who were craftsmen or operatives is statistically significant at the 10% level. We also see a statistically significant lower fraction of children in normal school counties with different occupations in 1860 than their parents, excluding sons with a non-occupational response. This suggests less occupational mobility for children growing up in normal school counties.

As Long and Ferrie (2013) highlight, differences in the fraction of children who have different occupations than their parents could be due to differences in the distribution of occupations in normal school and asylum counties, or differences in the association between father's and son's occupations. We do not implement the Long and Ferrie (2013) method for distinguishing these two explanations, given that we find less mobility in normal school counties, which goes against the hypothesis that greater social mobility today is explained by pre-existing greater levels of occupational mobility. However, we decompose this difference in mobility to understand what is driving the difference. Table A7 shows that it is driven by asylum counties having greater movement away from the father's occupation of craftsmen or operatives, and greater movement into farming among sons whose fathers were not farmers.

Table A6: Father and Son Occupations, Normal School and Asylum Counties

	(1)	(2)	(3)	(4)	(5)	(6)
	Unskilled	Craft/Oper	White Collar	Non Occ	Farmer	Diff from Father
Panel A: Father Occupations 1850, Fathers of 1850 16-18 Year-Old White Males						
Grew up in Normal School County	-0.008	-0.028^{+}	-0.019	0.005	0.050	
	(0.008)	(0.016)	(0.020)	(0.008)	(0.038)	
Observations	102	102	102	102	102	
R-Squared	0.360	0.340	0.106	0.281	0.337	
Mean DV, Asylum Counties	0.048	0.146	0.071	0.028	0.706	
p-value randomization inference	0.376	0.096	0.155	0.501	0.105	
Panel B: Son Occupations 1860, 1850 16-18 Year-Old White Males						
Grew up in Normal School County	-0.008	-0.010	-0.013	-0.006	0.035	-0.034*
	(0.013)	(0.013)	(0.015)	(0.015)	(0.023)	(0.017)
Observations	102	102	102	102	102	102
R-Squared	0.273	0.560	0.233	0.203	0.512	0.436
Mean DV, Asylum Counties	0.161	0.164	0.100	0.215	0.360	0.501
p-value randomization inference	0.582	0.444	0.265	0.747	0.126	0.091

Notes: $^{+}$ $p < 0.1$, * $p < .05$, ** $p < .01$ This table is similar to Table A5, but using only males 16-18 years-old and living with a parent in 1850. We show their fathers' occupations in 1850 (panel A), and their occupations in 1860 (panel B). See text and notes to Table A5 for details. Robust standard errors are in parentheses. P-values based on randomization inference are obtained by permuting within states the assignment of normal school counties, among our sample of normal school and asylum counties. This is based on 1000 permutations.

In addition to studying occupational mobility, we examine other 1860 outcomes conditional on parental socioeconomic status. These measures are more similar to the measures from Chetty et al. (2018), though as we will discuss the variables available in the 1850 and 1860 censuses that we use in this section have limitations. We test for differences in the

Table A7: Decomposition of Occupational Differences Between Father and Son Occupations

	(1)	(2)	(3)	(4)	(5)
	Unskilled	Craft/Oper	White Collar	Non Occ	Farmer
Panel A: Father Occupations 1850, and Son had a Different Occupation in 1860					
Grew up in Normal School County	-0.006	-0.027*	-0.009	0.006	0.001
	(0.007)	(0.011)	(0.008)	(0.008)	(0.021)
Observations	102	102	102	102	102
R-Squared	0.253	0.226	0.123	0.288	0.277
Mean DV, Asylum Counties	0.034	0.088	0.040	0.028	0.311
p-value randomization inference	0.47	0.009	0.191	0.465	0.968
Panel B: Son Occupations 1860, and Father had a Different Occupation in 1850					
Grew up in Normal School County	-0.006	-0.007	-0.006		-0.017^{+}
	(0.015)	(0.011)	(0.011)		(0.009)
Observations	102	102	102		102
R-Squared	0.216	0.534	0.283		0.208
Mean DV, Asylum Counties	0.187	0.146	0.097		0.071
p-value randomization inference	0.723	0.547	0.599		0.08

Notes: $^{+}$ $p < 0.1$, * $p < .05$, ** $p < .01$. This table decomposes the result in Table A6, panel B, column 6. The numerator in panel A is the number of fathers with the occupation whose sons have a different occupation, and the denominator is the number of fathers (equal to the number of sons). In panel B, the numerator is the number of sons in the occupation whose fathers have a different occupation, and the denominator is the number of sons (equal to the number of fathers). In both panels we exclude the sons who had a non-occupational response in 1860. Robust standard errors are in parentheses. P-values based on randomization inference are obtained by permuting within states the assignment of normal school counties, among our sample of normal school and asylum counties. This is based on 1000 permutations.

fraction enrolled in school at all in the previous year excluding Sunday or evening schools (the definition of enrollment for the 1860 census), fraction who list their occupation as a student, fraction married, fraction with top quartile household real estate value among White 25-28 year olds (400 dollars), and fraction with top quartile household personal estate value (372 dollars). We define household as the respondent and the respondent's spouse if they have one, using the IPUMS imputed family relationships. These wealth variables are the only available variables related to wealth or income in the 1860 census, and they both may differentially measure farmers' wealth relative to non-farmers who may have wage incomes but no real estate. For this reason, we focus on differences in the fraction with top quartile wealth, to avoid capturing poor farmers with very little real estate or personal wealth.[51]

First, Table A8 shows there are not statistically significant differences in the parental wealth distribution between normal school and asylum counties in 1850.

Among children growing up in lower socieconomic status families, there is no statistically significant difference between those growing up in normal school or asylum counties in school enrollment, marriage, or wealth (Table A9). For higher socioeconomic status children, the fraction with top quartile real estate is higher for those who grew up in normal school counties, though there is no difference in the fraction with top quartile personal estate value. We do not see this as strong evidence for pre-existing differential mobility in normal school counties, as real estate wealth may differentially reflect farmers' wealth (and there was a higher fraction of farmers in normal school counties). Also we only see this for the wealthier children, and not for children of poorer families, and only for real estate and not personal estate. Finally, we see evidence of less mobility in normal school counties when looking at occupational mobility.

Column 6 of Table A9 shows results when the dependent variable is the first principal component of the variables in columns 1 through 5. This is calculated separately in Panel A and Panel B, by implementing principal components analysis using the county-level averages of the variables in columns 1 through 5 among children in each panel. All variables enter this first component positively except fraction enrolled and fraction whose occupation is student. Among children from low socioeconomic status families, there is no significant difference for those from normal school versus asylum counties. The difference is significant at the 10% level for children from higher socioeconomic status families.

[51]The instructions for the value of personal estate specify that this should include all wealth not included in real estate wealth, which may include "the value of bonds, mortgages, notes, slaves, live stock, plate, jewels, or furniture" (Census Office, 1860).

Table A8: Parental Real Estate Wealth of 1850 16-18 Year-Old White Children

	Parental Real Estate Wealth, 1850		
	[0,150]	(150,1000]	> 1000
Grew up in Normal School County	0.005	0.028	-0.033
	(0.037)	(0.027)	(0.035)
Observations	102	102	102
R-Squared	0.236	0.469	0.335
Mean DV, Asylum Counties	0.356	0.270	0.374
p-value randomization inference	0.876	0.266	0.328

Notes: $^{+}$ $p < 0.1$, * $p < .05$, ** $p < .01$. Outcomes are county-level fraction with parental real estate wealth in each bin in 1850, among individuals 16-18 years old and living with at least one parent in 1850, who could be matched to their 1860 records using The Census Tree (Price et al., 2023*a*). We use the 1850-1990 county crosswalk from Eckert et al. (2020). There are 15 states in the regressions. Robust standard errors are in parentheses. P-values based on randomization inference are obtained by permuting within states the assignment of normal school counties, among our sample of normal school and asylum counties. This is based on 1000 permutations.

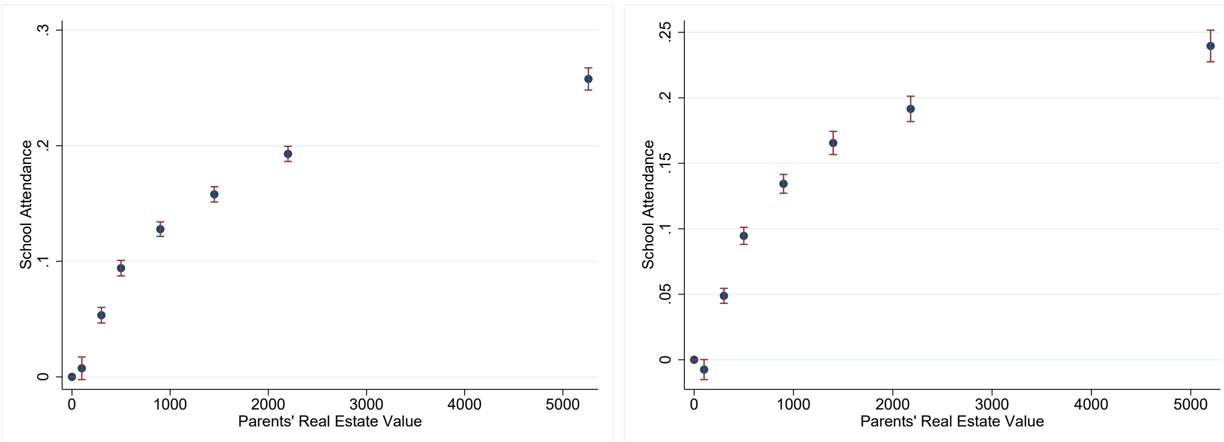
Table A9: Child Outcomes 1860, 16-18 Year-Old White Children in 1850

	(1)	(2)	(3)	(4)	(5)	(6)
	Enrolled	Student	Married	Top Quartile Real Estate	Top Quartile Personal Estate	PC 1
Panel A: Children with Parents' Real Estate Wealth in 1850 $\leq$ \$150						
Grew up in Normal School County	-0.002	-0.007	0.005	0.008	0.005	0.156
	(0.002)	(0.006)	(0.016)	(0.016)	(0.013)	(0.202)
Observations	102	102	102	102	102	102
R-Squared	0.202	0.276	0.404	0.510	0.548	0.657
Mean DV, Asylum Counties	0.010	0.014	0.593	0.234	0.203	0.189
p-value randomization inference	0.462	0.457	0.761	0.58	0.725	0.381
Panel B: Children with Parents' Real Estate Wealth in 1850 > \$150						
Grew up in Normal School County	-0.001	-0.005	0.020	0.028*	-0.002	0.373^{+}
	(0.002)	(0.005)	(0.017)	(0.013)	(0.016)	(0.212)
Observations	102	102	102	102	102	102
R-Squared	0.187	0.210	0.346	0.463	0.590	0.543
Mean DV, Asylum Counties	0.011	0.013	0.579	0.321	0.316	-0.138
p-value randomization inference	0.521	0.588	0.268	0.071	0.899	0.112

Notes: $^{+}$ $p < 0.1$, * $p < .05$, ** $p < .01$. Outcomes are county-level averages of 1860 outcomes for individuals who were 16-18 year old in 1850 and living with at least one parent in 1850, by parental real estate wealth in 1850. These averages are calculated among individuals who could be matched to their 1860 records using The Census Tree (Price et al., 2023*a*). The enrollment variable is based on the census question on school enrollment, and the student variable is based on the occupation question. PC 1 denotes the first principal component of the variables in columns 1 through 5. This is calculated separately in Panel A and Panel B, by implementing principal components analysis using the county-level averages of the variables in columns 1 through 5 among children in each panel. We use the 1850-1990 county crosswalk from Eckert et al. (2020). There are 15 states in the regressions. Robust standard errors are in parentheses. P-values based on randomization inference are obtained by permuting within states the assignment of normal school counties, among our sample of normal school and asylum counties. This is based on 1000 permutations.

B.2 Historical measures of educational mobility

In this section we show that the likelihood of school attendance in 1850 increases with parents' real estate value. This suggests that the fraction of children in the county attending school, among those with parents whose real estate value is less than or equal to 150 dollars, is reflective of the extent of upward mobility in the county. We use the same states as the exercise above, but do not link across censuses.



(a) 14-17 year olds

(b) 7-13 year olds

Figure A2: **School Attendance of White Children in 1850 by Parents' Real Estate Value, relative to those whose parents have real estate value of zero, with county fixed effects**. Estimates are from a regression of an indicator for school attendance on indicators for deciles of parents' real estate value, and including county fixed effects. Standard errors clustered at the county level. Sample includes White children who were living with at least one of their parents.

B.3 Differences in Economic Mobility in 1920-1940: Evidence from the Census Tree

The previous section showed there were no differences in economic and social mobility before the normal schools were established. We are also interested in whether the normal schools had impacts on mobility in the early or mid 20th century, when they had become teachers colleges or state colleges.

In order to obtain measures of mobility more similar to those we present using the Opportunity Insights data, we also use the Census Tree to link individuals across the 1920 and 1940 censuses. We identify six to fifteen year-olds in the 1920 full count census who are living in normal school or asylum counties, and living with at least one of their parents (Ruggles et al., 2021*a*), and not in group quarters. We then link these records to their record in the 1940 full count census when they should be 26 to 35 years old using the 1920 to 1940 Census Tree crosswalk (Price et al., 2023*b*) and the data from the 1940 full count of the census (Ruggles et al., 2021*a*).[52] We use these ages so that there is sufficient time to measure completion of high school and college in the 1940 census.

Of 2,969,684 White males ages 6-15 in the 1920 census who grew up in normal school or asylum counties, 2,035,528 (roughly 69%) have links between their 1920 and 1940 census records in the Census Tree.[53] Of 2,941,528 White females ages 6-15 in the 1920 census who grew up in normal school or asylum counties, 1,192,604 (roughly 41%) have links between their 1920 and 1940 census records.

There are no statistically significant differences between normal school and asylum counties in the fraction whose ID is linked to the 1920-1940 Census Tree crosswalk based on their 1920 ID, for men or for women.[54]

There is a positive relationship between merging to the crosswalk and socioeconomic status. Regressing an indicator for merging on parent's occupation score at the individual level, including 1920 county fixed effects, and clustering at the 1920 county level, yields a positive and statistically significant coefficient. This may be a sample of individuals among whom proximity to college would have an impact. Thus, if we observed the full sample of people growing up in normal school and asylum counties in 1920 linked to their 1940 record, the effect of proximity may be much smaller. This is important for comparing the size of these coefficients to effects in Figure 2 based on Opportunity Insights data. On the other hand, effects may be larger in this earlier period when geographic frictions are larger, and people in asylum counties would have been even less likely to travel for college. We further address issues of selection into linked census records in Section B.4.

The 1920 census does not have any measures of income or educational attainment, and so we construct several measures of mobility, similar to our analysis of 1850 children. First,

[52] When matching to the 1940 census records, we include individuals age 23 to 38.

[53] We note that some of these individuals do not have links to their 1940 record because they are no longer living in 1940.

[54] A small number of those who merge to the 1920 ID in the 1920-1940 crosswalk do not merge to a 1940 census record, presumably because we restricted the 1940 full count census file to those who were 23 to 38 in 1940 (to align with the ages we selected for the 1920 census). The link rates above only look at likelihood of merging to the 1920 ID in the crosswalk, which captures the vast majority of the lack of merge to the crosswalk. Of the 3,228,132 6-15 year-olds in 1920 who merge to the crosswalk, 33,417 (1%) do not have 1940 census records, presumably because of the age restriction we imposed on the 1940 file.

we look at differences in the occupation distribution, and the fraction of individuals with occupations different from their parents. Second, we use parental occupation as a way of identifying children from lower socioeconomic status families. Specifically, we use the 1950 occupational income score, acknowledging that occupational standing for occupations in 1950 may not reflect 1920 standing. Lower socioeconomic status individuals are defined as individuals whose parents' maximal occupation score in 1920 was less than or equal to the median for the states in our regression sample (the median score was 20).

For the sample of White 6-15 year olds in 1920 whose records linked to their 1940 census record and whose parents had maximal occupation score in 1920 less than or equal to the median for this sample, roughly 75% were in three occupation groups: farmers (owners and tenants); laborers (not elsewhere classified); and farm laborers, wage workers. Farmers (owners and tenants) make up 55% of the group with parental occupation score less than or equal to the median. Because farmers may be a heterogeneous group in terms of income, we also present results separating out the children of farmers.

We calculate county-level averages by race and sex, and examine several outcomes: completion of at least high school, at least some college, at least college, marital status, employment, and log household wage and salary income. We estimate a regression similar to equation (A1), but using the 1920 and 1940 censuses. We cluster standard errors at the 1920 state level, and we have 40 states in this analysis.

Tables A10 and A11 show the only statistically significant difference in the occupational distributions of parents or children are that children who grew up in normal school counties are more likely to be farmers in 1940 (significant at the 10% level), and males are more likely to have a non-occupational response in 1940. While not statistically significant, we also see that among children growing up in normal school counties, there is a smaller fraction with a different occupation than their parents. Thus, based on occupational measures, we do not see evidence of greater economic and social mobility in normal school counties when the normal schools had become colleges, but before their large increase in size. Tables A12 and A13 show similar results for Black children growing up in normal school and asylum counties. These are based on a smaller number of states given the sample restriction we described above, that states must have more than half of their normal school counties and more than half of their asylum counties with sample sizes of at least 10.

Panels A and B of Table A15 show the effects on other outcomes for men and women from lower socioeconomic status families. Among White males from lower socioeconomic status families, children growing up in normal school counties are .7 percentage points more likely to have completed high school and at least some college, though only the latter is significant at the 10% level. This latter effect is an increase of roughly 6%, based on the mean of the dependent variable in asylum counties. The magnitude of the effect on college completion is not statistically significant. There are also insignificant differences in the fraction married, and log average household income. There is a slight decrease in fraction employed, significant at the 10% level.

For White women from lower socioeconomic status families, those growing up in normal school counties are more likely to graduate from high school, though this is not statistically significant. The fraction with at least some college is 2.4 percentage points higher (roughly a 17% increase), and the fraction completing college is higher by .8 percentage points (also roughly a 17% increase). Both of these effects are statistically significant at the 1% level.

Table A10: Parent and Child Occupations, Normal School and Asylum Counties

	(1)	(2)	(3)	(4)	(5)	(6)
	Unskilled	Craft/Oper	White Collar	Non Occ	Farmer	Diff from Parent
Panel A: Parent Occupations 1920, Parents of 1920 6-15 Year-Old White Children						
Grew up in Normal School County	0.000	-0.016	-0.002	-0.004	0.025	
	(0.007)	(0.015)	(0.008)	(0.005)	(0.024)	
Observations	314	314	314	314	314	
R-Squared	0.401	0.552	0.256	0.172	0.514	
Mean DV, Asylum Counties	0.160	0.268	0.172	0.049	0.342	
Panel B: Child Occupations 1940, 1920 6-15 Year-Old White Children						
Grew up in Normal School County	0.004	-0.008	-0.012	0.002	0.013^{+}	-0.005
	(0.006)	(0.007)	(0.009)	(0.001)	(0.008)	(0.006)
Observations	314	314	314	314	314	314
R-Squared	0.342	0.551	0.341	0.205	0.584	0.342
Mean DV, Asylum Counties	0.200	0.342	0.353	0.054	0.104	0.660

Notes: $^{+}$ $p < 0.1$, * $p < .05$, ** $p < .01$. This table is the equivalent of Table A5, but for children in 1920, linked to their 1940 census record. We use the 1920-1990 county crosswalk from Eckert et al. (2020). There are 40 states in the regression. Standard errors clustered at the state level are in parentheses. See Table A5 and text for details.

Table A11: Father and Son Occupations, Normal School and Asylum Counties

	(1)	(2)	(3)	(4)	(5)	(6)
	Unskilled	Craft/Oper	White Collar	Non Occ	Farmer	Diff from Father
Panel A: Father Occupations 1920, Fathers of 1920 6-15 Year-Old White Males						
Grew up in Normal School County	-0.006	-0.014	-0.002	-0.002	0.025	
	(0.007)	(0.016)	(0.009)	(0.004)	(0.025)	
Observations	314	314	314	314	314	
R-Squared	0.470	0.539	0.248	0.170	0.516	
Mean DV, Asylum Counties	0.152	0.274	0.170	0.021	0.355	
Panel B: Son Occupations 1940, 1920 6-15 Year-Old White Males						
Grew up in Normal School County	0.003	-0.008	-0.010	0.003^{*}	0.012	-0.005
	(0.007)	(0.008)	(0.009)	(0.001)	(0.008)	(0.007)
Observations	314	314	314	314	314	314
R-Squared	0.339	0.545	0.326	0.119	0.581	0.351
Mean DV, Asylum Counties	0.208	0.355	0.294	0.036	0.107	0.628

Notes: $^{+}$ $p < 0.1$, * $p < .05$, ** $p < .01$. This table is the equivalent of Table A6, but for children in 1920, linked to their 1940 census record. We use the 1920-1990 county crosswalk from Eckert et al. (2020). There are 40 states in the regression. Standard errors clustered at the state level are in parentheses. See Table A6 and text for details.

Table A12: Parent and Child Occupations, Normal School and Asylum Counties

	(1)	(2)	(3)	(4)	(5)	(6)
	Unskilled	Craft/Oper	White Collar	Non Occ	Farmer	Diff from Parent
Panel A: Parent Occupations 1920, Parents of 1920 6-15 Year-Old Black Children						
Grew up in Normal School County	0.002	-0.005	-0.010	-0.006	0.016	
	(0.028)	(0.020)	(0.007)	(0.010)	(0.028)	
Observations	195	195	195	195	195	
R-Squared	0.256	0.357	0.216	0.102	0.662	
Mean DV, Asylum Counties	0.594	0.193	0.053	0.040	0.168	
p-value randomization inference	0.94	0.81	0.135	0.473	0.585	
Panel B: Child Occupations 1940, 1920 6-15 Year-Old Black Children						
Grew up in Normal School County	-0.025	-0.015	0.003	0.017	0.018^{+}	0.003
	(0.018)	(0.015)	(0.012)	(0.012)	(0.009)	(0.022)
Observations	195	195	195	195	195	195
R-Squared	0.231	0.165	0.247	0.132	0.575	0.202
Mean DV, Asylum Counties	0.577	0.207	0.119	0.111	0.042	0.539
p-value randomization inference	0.161	0.414	0.785	0.19	0.059	0.888

Notes: $^{+}$ $p < 0.1$, * $p < .05$, ** $p < .01$. This table is the equivalent of Table A5, but for children in 1920, linked to their 1940 census record. We use the 1920-1990 county crosswalk from Eckert et al. (2020). Standard errors robust to heteroskedasticity are in parentheses. P-values based on randomization inference are obtained by permuting within states the assignment of normal school counties, among our sample of normal school and asylum counties. This is based on 1000 permutations. See Table A5 and text for details.

Table A13: Father and Son Occupations, Normal School and Asylum Counties

	(1)	(2)	(3)	(4)	(5)	(6)
	Unskilled	Craft/Oper	White Collar	Non Occ	Farmer	Diff from Father
Panel A: Father Occupations 1920, Fathers of 1920 6-15 Year-Old Black Males						
Grew up in Normal School	-0.012	0.010	-0.005	-0.010	0.024	
	(0.036)	(0.025)	(0.009)	(0.010)	(0.035)	
Observations	164	164	164	164	164	
R-Squared	0.314	0.347	0.208	0.120	0.618	
Mean DV, Asylum Counties	0.498	0.175	0.044	0.022	0.235	
p-value randomization inference	0.774	0.738	0.513	0.177	0.498	
Panel B: Son Occupations 1940, 1920 6-15 Year-Old Black Males						
Grew up in Normal School	-0.017	-0.013	0.012	-0.007	0.024^{+}	-0.015
	(0.024)	(0.022)	(0.015)	(0.013)	(0.012)	(0.026)
Observations	164	164	164	164	164	164
R-Squared	0.130	0.126	0.248	0.103	0.486	0.212
Mean DV, Asylum Counties	0.533	0.230	0.099	0.082	0.056	0.583
p-value randomization inference	0.535	0.63	0.444	0.598	0.092	0.598

Notes: $^{+}$ $p < 0.1$, * $p < .05$, ** $p < .01$. This table is the equivalent of Table A6, but for children in 1920, linked to their 1940 census record. We use the 1920-1990 county crosswalk from Eckert et al. (2020). Standard errors robust to heteroskedasticity are in parentheses. P-values based on randomization inference are obtained by permuting within states the assignment of normal school counties, among our sample of normal school and asylum counties. This is based on 1000 permutations. See Table A6 and text for details.

Table A14: Work Relief: Unemployed for at least 65 weeks, and worked at least 40 weeks in 1939

	(1)	(2)	(3)	(4)
Panel A: White children				
Grew up in Normal School County	0.002*	0.003**	0.002	0.0002
	(0.001)	(0.001)	(0.001)	(0.001)
Observations	314	314	314	314
R-squared	0.241	0.244	0.226	0.198
Mean DV, Asylum Counties	0.006	0.003	0.009	0.005
Sex	Male	Female	Male	Female
Parental SES	High	High	Low	Low
Panel B: Black children				
Grew up in Normal School County	0.008	0.001	-0.003	0.001
	(0.009)	(0.005)	(0.007)	(0.005)
Observations	89	49	142	76
R-squared	0.151	0.065	0.132	0.140
Mean DV, Asylum Counties	0.012	0.004	0.030	0.006
Sex	Male	Female	Male	Female
Parental SES	High	High	Low	Low
p-value, randomization inference	0.265	0.972	0.736	0.942

Notes: $^{+}$ $p < 0.1$, * $p < .05$, ** $p < .01$. Dependent variable is a measure of work relief: the fraction of individuals who reported working at least 40 weeks in 1939, and also report being unemployed for at least 65 weeks on the day of the census (April 1, 1940), among those who reported working at least 40 weeks in 1939. Standard errors in panel A are clustered at the state level. In panel B, standard errors robust to heteroskedasticity are in parentheses. P-values based on randomization inference are obtained by permuting within states the assignment of normal school counties, among our sample of normal school and asylum counties. This is based on 1000 permutations. See notes to Table 2 for details.

They are 1.9 percentage points more likely to be married (roughly 3%), and 2 percentage points less likely to be employed (roughly 7%), and there is no significant difference in household income.

Panels B and C show the effects for higher socioeconomic status children. The effects of growing up in a normal school county on male educational attainment are much larger in magnitude, and also larger in percentage terms than the effects for lower socioeconomic status children. Despite these education effects, there are not differences in household wage and salary income. While college attainment is higher, local colleges for this sample of children were still focused on training teachers, rather than higher-wage occupations, and this may explain the absence of an effect on income. Further, in 1940, individuals are asked only for their wage and salary income as an employee, which may downward-bias the income effects in our setting because of the greater fraction of farmers among children growing up in normal school counties.

Similarly, the effects of growing up in a normal school county on educational attainment of higher socioeconomic status females are much larger in magnitude and also larger in relative terms than the effects for lower socioeconomic status females. While the effects on marriage, employment, and household income are similar in size, these are not statistically significant, except the employment effects at the 10% level. Table A17 shows that excluding the children of farmers from Table A15 yields similar results.

Table A16 shows results for Black children, which we do not emphasize as they are based on a smaller set of states for the sample size reasons discussed above. Most of the effects are not statistically significant, although we see that for lower socioeconomic status males, they are more likely to have some college and more likely married. Many southern states established normal schools for Black students (which later became HBCUs), and there is also evidence that Black students enrolled at Northern institutions that had been normal schools (Ogren, 2003). These both could explain the impacts on education for Black children in this earlier period.

We note that in the 1940 census, wage and salary income reported for 1939 should include income on work relief. This is true even if the respondent was in a non-relief job in 1940. As an additional measure of labor market outcomes, we construct a measure of work relief in 1939, and test whether children growing up in normal school counties were differentially likely to be on work relief as adults. Specifically, we identify people who report being unemployed for 65 weeks or more on the day of the census (April 1, 1940), but reported working at least 40 weeks in 1939. We follow the same methodology as for the other variables in Table 2, and construct the county-level average among people who reported working at least 40 weeks in 1939. Based on this measure, the evidence is consistent with higher rates of work relief in normal school counties in 1939, especially for children from higher socioeconomic status families (Table A14). This is consistent with the large number of work relief projects that took place on college campuses (Living New Deal, 2024).

These results suggest the teachers colleges and the state colleges that the normal schools had evolved into by the 1930s were already affecting access to education in their local communities. Our results using the Opportunity Insights data shows this process has continued to the present day, including for children from very poor families whose parents likely did not attain a college education because of the previous normal school. The effects appear larger in percentage terms than the effects in more recent years using the Opportunity Insights data.

This may reflect greater geographic frictions in college access in the early 1900s. Importantly, we do not see any differences in educational mobility in 1850 before the normal schools had been established in most places (Table A1), suggesting these educational institutions are the channel.

Table A15: 1940 Outcomes of 1920 6-15 Year-Old White Children, County-Level Outcomes based on 1920 County

	(1)	(2)	(3)	(4)	(5)	(6)
	$\geq$ HS	$\geq$ Some College	$\geq$ College	Married	Employed	Ln(HH Income)
Panel A: Males, Lower Socioeconomic Status						
Grew up in Normal School County	0.007	0.007^{+}	0.002	0.005	-0.006^{+}	-0.026
	(0.007)	(0.004)	(0.002)	(0.004)	(0.003)	(0.016)
Observations	314	314	314	314	314	314
R-Squared	0.583	0.497	0.425	0.640	0.362	0.495
Mean DV, Asylum Counties	0.308	0.113	0.055	0.707	0.901	7.027
Panel B: Females, Lower Socioeconomic Status						
Grew up in Normal School County	0.010	0.024**	0.008**	0.019*	-0.020**	-0.013
	(0.010)	(0.006)	(0.003)	(0.008)	(0.007)	(0.015)
Observations	314	314	314	314	314	314
R-Squared	0.474	0.408	0.355	0.675	0.657	0.471
Mean DV, Asylum Counties	0.404	0.140	0.047	0.701	0.297	7.052
Panel C: Males, Higher Socioeconomic Status						
Grew up in Normal School County	0.016*	0.025**	0.013**	0.001	-0.003	0.001
	(0.006)	(0.005)	(0.004)	(0.005)	(0.003)	(0.012)
Observations	314	314	314	314	314	314
R-Squared	0.588	0.565	0.372	0.529	0.192	0.378
Mean DV, Asylum Counties	0.502	0.236	0.126	0.705	0.897	7.233
Panel D: Females, Higher Socioeconomic Status						
Grew up in Normal School County	0.021*	0.045**	0.025**	0.012	-0.013^{+}	-0.005
	(0.008)	(0.008)	(0.005)	(0.007)	(0.007)	(0.010)
Observations	314	314	314	314	314	314
R-Squared	0.521	0.535	0.442	0.697	0.636	0.423
Mean DV, Asylum Counties	0.576	0.242	0.106	0.650	0.365	7.229

Notes: $^{+}$ $p < 0.1$, * $p < .05$, ** $p < .01$. Outcomes are county-level averages, separately for those with lower and higher socioeconomic status in 1920, among individuals who could be matched to their 1940 records using The Census Tree (Price et al., 2023*b*). Lower socioeconomic status individuals are defined as individuals whose parents' maximal occupation score in 1920 was less than or equal to the median for the states in our regression sample (the median score was 20). The variable Married is the fraction of individuals who are married with spouse present. The variable Ln(HH Income) is the log of the county-level average wage and salary income of individuals and their spouses. The county-level average is constructed by taking the total wage and salary income of individuals and their spouses and dividing by the total number of individuals (males in Panels A and C and females in Panels B and D) with positive own or spousal wage and salary income. We use the 1920-1990 county crosswalk from Eckert et al. (2020). Standard errors are clustered at the 1920 state level.

Table A16: 1940 Outcomes of 1920 6-15 Year-Old Black Children, County-Level Outcomes based on 1920 County

	(1)	(2)	(3)	(4)	(5)	(6)
	$\geq$ HS	$\geq$ Some College	$\geq$ College	Married	Employed	Ln(HH Income)
Panel A: Males, Lower Socioeconomic Status						
Grew up in Normal School County	0.023	0.039*	0.002	0.033^{+}	0.010	0.066
	(0.019)	(0.018)	(0.004)	(0.019)	(0.015)	(0.051)
Observations	142	142	142	142	142	142
R-Squared	0.380	0.168	0.162	0.362	0.242	0.330
Mean DV, Asylum Counties	0.181	0.054	0.020	0.577	0.809	6.446
p-value randomization inference	0.314	0.049	0.553	0.118	0.518	0.206
Panel B: Females, Lower Socioeconomic Status						
Grew up in Normal School County	-0.020	0.015	0.000	0.027	-0.012	0.009
	(0.021)	(0.011)	(0.009)	(0.020)	(0.025)	(0.053)
Observations	77	77	77	77	77	77
R-Squared	0.431	0.401	0.253	0.524	0.213	0.428
Mean DV, Asylum Counties	0.196	0.080	0.038	0.324	0.515	6.045
p-value randomization inference	0.367	0.259	0.988	0.216	0.648	0.876
Panel C: Males, Higher Socioeconomic Status						
Grew up in Normal School County	0.026	0.042	0.035	-0.015	0.020	-0.140
	(0.038)	(0.030)	(0.030)	(0.035)	(0.027)	(0.086)
Observations	93	93	93	93	93	93
R-Squared	0.194	0.255	0.161	0.136	0.118	0.112
Mean DV, Asylum Counties	0.279	0.123	0.054	0.595	0.809	6.641
p-value randomization inference	0.582	0.252	0.36	0.727	0.561	0.27
Panel D: Females, Higher Socioeconomic Status						
Grew up in Normal School County	-0.007	-0.011	0.018	0.024	-0.032	-0.149
	(0.051)	(0.045)	(0.028)	(0.037)	(0.055)	(0.128)
Observations	50	50	50	50	50	49
R-Squared	0.240	0.328	0.348	0.240	0.164	0.190
Mean DV, Asylum Counties	0.362	0.182	0.076	0.227	0.568	6.359
p-value randomization inference	0.915	0.835	0.584	0.617	0.585	0.229

Notes: $^{+}$ $p < 0.1$, * $p < .05$, ** $p < .01$. This table is the same as Table A15, but for Black children. Standard errors robust to heteroskedasticity are in parentheses. P-values based on randomization inference are obtained by permuting within states the assignment of normal school counties, among our sample of normal school and asylum counties. This is based on 1000 permutations.

Table A17: 1940 Outcomes of 1920 6-15 Year-Old White Children, County-Level Outcomes based on 1920 County, Excluding Children of Farmers

	(1)	(2)	(3)	(4)	(5)	(6)
	$\geq$ HS	$\geq$ Some College	$\geq$ College	Married	Employed	Ln(HH Income)
Panel A: Males, Lower Socioeconomic Status						
Grew up in Normal School County	0.002	0.007	0.002	0.004	-0.004	-0.027
	(0.007)	(0.005)	(0.003)	(0.005)	(0.004)	(0.018)
Observations	314	314	314	314	314	314
R-Squared	0.515	0.400	0.297	0.524	0.251	0.332
Mean DV, Asylum Counties	0.327	0.129	0.064	0.697	0.880	7.068
Panel B: Females, Lower Socioeconomic Status						
Grew up in Normal School County	0.007	0.025**	0.008*	0.018*	-0.017*	-0.011
	(0.008)	(0.006)	(0.003)	(0.008)	(0.008)	(0.016)
Observations	314	314	314	314	314	314
R-Squared	0.444	0.391	0.305	0.638	0.589	0.341
Mean DV, Asylum Counties	0.406	0.139	0.055	0.683	0.323	7.093

Notes: $^{+}$ $p < 0.1$, * $p < .05$, ** $p < .01$. This table is similar to Panels A and B of Table A15, but excludes the children whose parents' occupation, for the parent who had the maximal occupation score, is farmer (owners and tenants) or farm manager. These children are included in Table A15 because the occupation score for these occupations is below the median. See notes to Table A15 and text for details. We use the 1920-1990 county crosswalk from Eckert et al. (2020). Standard errors are clustered at the 1920 state level.

B.4 Representativeness of Census Tree Links

In this section, we follow the recommendations of Bailey, Cole and Massey (2020) for addressing the possibility of non-representative samples in linked data. First we estimate regressions in which the dependent variable is an indicator for merging to the Census Tree, and the covariates include various demographic, economic, and family characteristics. We then estimate the propensity to be matched, and use those to weight the observations. We do this for the 1850-1860 analyses as well as the 1920-1940 analyses. We describe this in detail below.

For the 1850-1860 analysis, we test whether the set of 16-18 year-olds in 1850 that match to their records in 1860 are representative of all 16-18 year-olds in the 1850 census in our sample of counties. We estimate a regression in which the dependent variable is an indicator for whether the individual is in our 1850 and 1860 data, and the individual's record is linked. We include the following independent variables from the 1850 census: male, maximum occupation score of the individual's parents, mother's age, father's age, whether the father is the head of the household, whether the mother is the head of the household, whether the individual lives with their mother, whether the individual lives with their father, whether the individual lives in their birth state, whether the father lives in his birth state, whether the mother lives in her birth state, whether the individual is foreign born, whether the individual's father is foreign born, whether the individual's mother is foreign born, the number of siblings, and the family unit in the household to which the person belongs.

Similarly, for the 1920-1940 analysis, we test whether the set of 6-15 year-olds in 1920 that match to their records in 1940 are representative of all 6-15 year-olds in the 1920 census in our sample of counties. We estimate a regression in which the dependent variable is an indicator for whether the individual is in our 1920 and 1940 data, and the individual's record is linked. We include the same independent variables from the 1920 census as we did for the 1850-1860 analysis, except male since all of our analysis in 1920 is by gender.

Given that our specifications in Appendix B.1 and B.3 are estimated separately by race, sex, and socioeconomic status, we also estimate these regressions separately by the same subgroups as in our main analysis. This allows us to test for differences in selection, even within the subgroups of interest. We weight the observations by the weights from the Eckert et al. (2020) county crosswalks, and include county fixed effects. We cluster standard errors at the county level. We include the same states that were included in our main results in Table 1 and 2.

Across all of our subgroups we see that there is selection into being in the Census Tree linked sample. Table A18 shows many variables that are positively correlated with linking of 1850 and 1860 records, and Tables A24 and A25 show this for the linking of the 1920 and 1940 censuses. For example, linking is generally positively correlated with the individual living in their birth state in 1850 or 1920 when we observe them as children, and with the number of siblings, and negatively correlated with parents being born outside the U.S. Unless there are differences across normal school and asylum counties in selection into linked census records, our results in Tables 1 and 2 will not be biased for the selected subsample they pertain to, but they will be an estimate that is reflective of mobility for this particular subsample, and not reflective of overall mobility. The results here also suggest the individuals in the linked sample may be less geographically mobile, and thus the effects of college proximity may be particularly important for this sample. As a result, the effects may be smaller if we had all individuals.

Because of this selection into having a linked census record, we estimate the regressions for Table 1 and A15 including weights for representativeness based on propensity score estimation, following Bailey, Cole and Massey (2020). We first estimate the same regressions described above by subgroup, using a probit regression, and restricting to the same set of states and counties as in the main analysis.[55]

We get the predicted value of the Census Tree link variable, and construct the propensity score weight as one divided by this prediction. The total weight for each observation is the propensity score weight multiplied by the county crosswalk weight from Eckert et al. (2020). We use these weights to construct county-level averages of our dependent variables, and estimate the regressions in Tables 1 and 2, and Appendices B.1 and B.3. The results are generally quite similar, consistent with selection into linked census records that is not different across normal school and asylum counties within the same state, and similar treatment effects of normal school assignment in the merged sample and a representative sample.

[55]The main results dropped states in which at least half of the normal school or asylum counties in the state had sample sizes less than 10. To facilitate interpretation, we do not include these originally-dropped states, even if the weighted sample sizes increase above 10 for enough of the counties in the state.

Table A18: Testing for Representativeness of the Linked Census Tree Sample, White Individuals, 1850-1860

DV = Linked Sample	(1)	(2)
Male	0.183***	0.147***
	(0.010)	(0.007)
Parental occupation score	0.000	-0.001***
	(0.000)	(0.000)
Mother's age	0.002***	0.002***
	(0.000)	(0.000)
Father's age	-0.000	0.000
	(0.000)	(0.000)
Father head	-0.026	0.018
	(0.022)	(0.044)
Mother head	-0.043***	-0.068
	(0.013)	(0.050)
Lives with father	0.039	-0.043
	(0.027)	(0.049)
Lives with mother	-0.077***	-0.045*
	(0.020)	(0.024)
Lives in birth state	0.047***	0.046***
	(0.010)	(0.010)
Father lives in birth state	0.019***	0.006
	(0.007)	(0.006)
Mother lives in birth state	0.002	0.007
	(0.007)	(0.007)
Foreign born	0.066***	-0.013
	(0.014)	(0.019)
Father foreign born	-0.033**	-0.055***
	(0.013)	(0.016)
Mother foreign born	-0.052***	-0.034**
	(0.011)	(0.017)
Number of siblings	0.012***	0.008***
	(0.001)	(0.001)
First family unit in household	0.097***	0.113
	(0.017)	(0.069)
Sample	Low SES	High SES
Observations	51,164	80,369
R-squared	0.071	0.055
DV mean	0.424	0.544

Notes: $^{+}$ $p < 0.1$, * $p < .05$, ** $p < .01$. The outcome is whether the individual is in our 1850 and 1860 census data, and the record is linked through The Census Tree (Price et al., 2023*a*). Regressions include county fixed effects, and standard errors are clustered at the county level. See text for details.

Table A19: Economic and Social Mobility for Children in 1850, with Propensity Score Weights

	(1)	(2)	(3)	(4)	(5)	(6)
	Enrolled	Enrolled	Top Quartile Real Estate	Top Quartile Personal Estate	Married	Occ. Mobility
Year of Observation	1850	1850	1860	1860	1860	1860
Age in 1850	7-13	14-17	16-18	16-18	16-18	16-18
Grew up in Normal School County	-0.032	-0.005	0.004	0.004	0.003	-0.035**
	(.044)	(.037)	(0.016)	(0.013)	(0.016)	(0.017)
Observations	102	102	102	102	102	102
R-Squared	0.507	0.484	0.505	0.533	0.425	0.427
Mean DV, Asylum Counties	0.486	0.378	0.247	0.209	0.599	0.504
p-value randomization inference	0.474	0.899	0.809	0.775	0.851	0.078
Parental SES	Low	Low	Low	Low	Low	All

Notes: $^{+}$ $p < 0.1$, * $p < .05$, ** $p < .01$. This table is the equivalent of Table 1, but county-level averages are constructed from weighted individual-level observations, to increase representativeness of the sample. See text for details.

Table A20: Parent and Child Occupations, Normal School and Asylum Counties, with Propensity Score Weights

	(1)	(2)	(3)	(4)	(5)	(6)
	Unskilled	Craft/Oper	White Collar	Non Occ	Farmer	Diff from Parent
Panel A: Parent Occupations 1850, Parents of 1850 16-18 Year-Old White Children						
Grew up in Normal School County	-0.008	-0.019	-0.015	0.006	0.037	
	(0.006)	(0.013)	(0.014)	(0.013)	(0.034)	
Observations	102	102	102	102	102	
R-Squared	0.418	0.384	0.142	0.447	0.268	
Mean DV, Asylum Counties	0.042	0.125	0.071	0.168	0.594	
p-value randomization inference	0.228	0.13	0.128	0.632	0.206	
Panel B: Child Occupations 1860, 1850 16-18 Year-Old White Children						
Grew up in Normal School County	0.013	-0.010	-0.009	-0.010	0.027	-0.010
	(0.012)	(0.011)	(0.011)	(0.013)	(0.021)	(0.014)
Observations	102	102	102	102	102	102
R-Squared	0.257	0.549	0.333	0.287	0.538	0.410
Mean DV, Asylum Counties	0.136	0.155	0.094	0.295	0.337	0.578
p-value randomization inference	0.372	0.38	0.306	0.493	0.175	0.569

Notes: $^{+}$ $p < 0.1$, * $p < .05$, ** $p < .01$ This table is the equivalent of Table A5, but county-level averages are constructed from weighted individual-level observations, to increase representativeness of the sample. See text for details.

Table A21: Father and Son Occupations, Normal School and Asylum Counties, with Propensity Score Weights

	(1)	(2)	(3)	(4)	(5)	(6)
	Unskilled	Craft/Oper	White Collar	Non Occ	Farmer	Diff from Father
Panel A: Father Occupations 1850, Fathers of 1850 16-18 Year-Old White Males						
Grew up in Normal School County	-0.009	-0.030*	-0.019	0.005	0.053	
	(0.008)	(0.016)	(0.021)	(0.008)	(0.039)	
Observations	102	102	102	102	102	
R-Squared	0.362	0.335	0.104	0.289	0.334	
Mean DV, Asylum Counties	0.050	0.153	0.077	0.027	0.694	
p-value randomization inference	0.36	0.075	0.172	0.511	0.093	
Panel B: Son Occupations 1860, 1850 16-18 Year-Old White Males						
Grew up in Normal School County	-0.009	-0.009	-0.013	-0.005	0.036	-0.035**
	(0.013)	(0.013)	(0.015)	(0.015)	(0.023)	(0.017)
Observations	102	102	102	102	102	102
R-Squared	0.281	0.558	0.234	0.194	0.508	0.427
Mean DV, Asylum Counties	0.161	0.167	0.102	0.214	0.356	0.504
p-value randomization inference	0.533	0.467	0.248	0.755	0.122	0.078

Notes: $^{+}$ $p < 0.1$, * $p < .05$, ** $p < .01$ This table is the equivalent of Table A6, but county-level averages are constructed from weighted individual-level observations, to increase representativeness of the sample. See text for details.

Table A22: Parental Real Estate Wealth of 1850 16-18 Year-Old White Children, with Propensity Score Weights

	Parental Real Estate Wealth, 1850		
	[0,150]	(150,1000]	> 1000
Grew up in Normal School County	0.006	0.029	-0.035
	(0.037)	(0.027)	(0.034)
Observations	102	102	102
R-Squared	0.230	0.469	0.328
Mean DV, Asylum Counties	0.369	0.267	0.364
p-value randomization inference	0.862	0.242	0.302

Notes: $^{+}$ $p < 0.1$, * $p < .05$, ** $p < .01$. This table is the equivalent of Table A8, but county-level averages are constructed from weighted individual-level observations, to increase representativeness of the sample. See text for details.

Table A23: Child Outcomes 1860, 16-18 Year-Old White Children in 1850, with Propensity Score Weights

	(1)	(2)	(3)	(4)	(5)	(6)
	Enrolled	Student	Married	Top Quartile Real Estate	Top Quartile Personal Estate	PC 1
Panel A: Children with Parents' Real Estate Wealth in 1850 $\leq$ \$150						
Grew up in Normal School County	-0.003	-0.009	0.003	0.004	0.004	0.102
	(0.002)	(0.008)	(0.016)	(0.016)	(0.013)	(0.210)
Observations	102	102	102	102	102	102
R-Squared	0.223	0.275	0.425	0.505	0.533	0.641
Mean DV, Asylum Counties	0.010	0.015	0.599	0.247	0.209	0.231
p-value randomization inference	0.255	0.469	0.851	0.809	0.775	0.623
Panel B: Children with Parents' Real Estate Wealth in 1850 > \$150						
Grew up in Normal School County	-0.001	-0.006	0.020	0.026**	-0.004	0.340
	(0.002)	(0.006)	(0.017)	(0.013)	(0.016)	(0.218)
Observations	102	102	102	102	102	102
R-Squared	0.205	0.217	0.347	0.448	0.597	0.531
Mean DV, Asylum Counties	0.010	0.013	0.581	0.326	0.319	-0.123
p-value randomization inference	0.593	0.608	0.289	0.084	0.813	0.163

Notes: $^{+}$ $p < 0.1$, * $p < .05$, ** $p < .01$. This table is the equivalent of Table A9, but county-level averages are constructed from weighted individual-level observations, to increase representativeness of the sample. See text for details.

Table A24: Testing for Representativeness of the Linked Census Tree Sample, White Individuals, 1920-1940

	(1)	(2)	(3)	(4)
Parental occupation score	-0.001***	0.001***	-0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Mother's age	0.002***	0.002***	0.002***	0.003***
	(0.000)	(0.000)	(0.000)	(0.000)
Father's age	-0.000**	-0.000***	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Father head	0.023***	0.018***	0.027***	0.024***
	(0.005)	(0.003)	(0.005)	(0.004)
Mother head	0.018***	-0.010*	0.011***	0.005
	(0.003)	(0.006)	(0.003)	(0.004)
Lives with father	0.024***	0.034***	0.011*	0.040***
	(0.006)	(0.007)	(0.007)	(0.007)
Lives with mother	-0.061***	-0.050***	-0.073***	-0.076***
	(0.006)	(0.006)	(0.010)	(0.011)
Lives in birth state	0.048***	0.047***	0.035***	0.036***
	(0.004)	(0.002)	(0.003)	(0.003)
Father lives in birth state	0.018***	0.015***	0.019***	0.013***
	(0.003)	(0.002)	(0.004)	(0.003)
Mother lives in birth state	0.016***	0.010***	0.010**	0.007**
	(0.003)	(0.001)	(0.004)	(0.003)
Foreign born	-0.061***	-0.038***	-0.078***	-0.066***
	(0.008)	(0.006)	(0.007)	(0.003)
Father foreign born	-0.050***	-0.039***	-0.066***	-0.052***
	(0.003)	(0.002)	(0.004)	(0.003)
Mother foreign born	-0.059***	-0.050***	-0.076***	-0.062***
	(0.004)	(0.002)	(0.005)	(0.003)
Number of siblings	0.005***	0.002***	0.005***	0.000
	(0.000)	(0.000)	(0.000)	(0.001)
First family unit in household	0.111***	0.115***	0.097***	0.099***
	(0.007)	(0.007)	(0.005)	(0.006)
Sample	Low SES Males	High SES Males	Low SES Females	High SES Females
Observations	1,278,770	1,934,526	1,256,846	1,925,246
R-squared	0.040	0.026	0.076	0.050
DV mean	0.666	0.681	0.398	0.390

Notes: $^{+}$ $p < 0.1$, * $p < .05$, ** $p < .01$. The outcome is whether the individual is in the both the 1920 and 1940 census data, and the individual's record is linked through The Census Tree (Price et al., 2023*b*). Regressions include county fixed effects and standard errors are clustered at the county level. See text for details.

Table A25: Testing for Representativeness of Linked Census Tree Sample, Black Individuals, 1920-1940

	(1)	(2)	(3)	(4)
Parental occupation score	0.001	0.002***	0.000	0.002*
	(0.000)	(0.001)	(0.000)	(0.001)
Mother's age	0.000	0.001	0.000	0.000
	(0.000)	(0.001)	(0.000)	(0.001)
Father's age	-0.001***	-0.002***	-0.001***	-0.001*
	(0.000)	(0.001)	(0.000)	(0.001)
Father head	0.023**	-0.015	0.017*	0.020
	(0.011)	(0.026)	(0.009)	(0.020)
Mother head	-0.001	-0.031	-0.002	-0.056
	(0.010)	(0.037)	(0.006)	(0.039)
Lives with father	-0.043	0.071	0.066*	-0.017
	(0.028)	(0.063)	(0.037)	(0.057)
Lives with mother	0.078*	-0.016	-0.021	-0.044
	(0.042)	(0.068)	(0.050)	(0.082)
Lives in birth state	0.061***	0.063***	0.035***	0.029*
	(0.011)	(0.015)	(0.009)	(0.015)
Father lives in birth state	0.014***	0.026*	0.011***	0.033**
	(0.004)	(0.014)	(0.004)	(0.014)
Mother lives in birth state	-0.002	0.023	0.004	0.017
	(0.005)	(0.015)	(0.004)	(0.014)
Foreign born	-0.110*	0.089	0.067	-0.014
	(0.056)	(0.100)	(0.075)	(0.077)
Father foreign born	-0.115***	-0.037	0.033	0.001
	(0.029)	(0.041)	(0.035)	(0.039)
Mother foreign born	0.040	-0.020	-0.045	-0.047
	(0.042)	(0.062)	(0.049)	(0.068)
Number of siblings	0.013***	0.014***	0.006***	0.002
	(0.001)	(0.002)	(0.001)	(0.003)
First family unit in household	0.049***	0.043	0.031***	0.053
	(0.014)	(0.049)	(0.011)	(0.056)
Constant	0.248***	0.234***	0.003	0.085
	(0.037)	(0.059)	(0.066)	(0.119)
Sample	Low SES Males	High SES Males	Low SES Females	High SES Females
Observations	122,076	12,940	110,607	8,622
R-squared	0.019	0.021	0.021	0.021
DV mean	0.418	0.457	0.126	0.172

Notes: $^{+}$ $p < 0.1$, * $p < .05$, ** $p < .01$. The outcome is whether the individual is in both the 1920 and 1940 census data, and the individual's record is linked through The Census Tree (Price et al., 2023*b*). See text and Table A24 for details.

Table A26: Parent and Child Occupations, Normal School and Asylum Counties, with Propensity Score Weights

	(1) Unskilled	(2) Craft/Oper	(3) White Collar	(4) Non Occ	(5) Farmer	(6) Diff from Parent
Panel A: Parent Occupations 1920, Parents of 1920 6-15 Year-Old White Children						
Grew up in Normal School County	-0.000	-0.015	-0.002	-0.004	0.026	
	(0.007)	(0.015)	(0.008)	(0.005)	(0.024)	
Observations	314	314	314	314	314	
R-Squared	0.408	0.549	0.254	0.168	0.514	
Mean DV, Asylum Counties	0.164	0.268	0.168	0.054	0.337	
Panel B: Child Occupations 1940, 1920 6-15 Year-Old White Children						
Grew up in Normal School County	0.004	-0.008	-0.012	0.002	0.013^{+}	-0.005
	(0.006)	(0.007)	(0.009)	(0.001)	(0.008)	(0.006)
Observations	314	314	314	314	314	314
R-Squared	0.338	0.555	0.330	0.207	0.584	0.342
Mean DV, Asylum Counties	0.201	0.344	0.351	0.054	0.103	0.662

Notes: $^{+}$ $p < 0.1$, * $p < .05$, ** $p < .01$. This table is the equivalent of Table A10, but county-level averages are constructed from weighted individual-level observations, to increase representativeness of the sample. See text for details.

Table A27: Father and Son Occupations, Normal School and Asylum Counties, with Propensity Score Weights

	(1) Unskilled	(2) Craft/Oper	(3) White Collar	(4) Non Occ	(5) Farmer	(6) Diff from Father
Panel A: Father Occupations 1920, Fathers of 1920 6-15 Year-Old White Males						
Grew up in Normal School County	-0.007	-0.014	-0.002	-0.003	0.025	
	(0.007)	(0.016)	(0.008)	(0.005)	(0.025)	
Observations	314	314	314	314	314	
R-Squared	0.472	0.535	0.248	0.170	0.516	
Mean DV, Asylum Counties	0.155	0.275	0.167	0.022	0.354	
Panel B: Son Occupations 1940, 1920 6-15 Year-Old White Males						
Grew up in Normal School County	0.003	-0.008	-0.010	0.003*	0.012	-0.005
	(0.007)	(0.008)	(0.009)	(0.001)	(0.008)	(0.007)
Observations	314	314	314	314	314	314
R-Squared	0.336	0.546	0.320	0.120	0.582	0.347
Mean DV, Asylum Counties	0.209	0.356	0.293	0.036	0.107	0.629

Notes: $^{+}$ $p < 0.1$, * $p < .05$, ** $p < .01$. This table is the equivalent of Table A11, but county-level averages are constructed from weighted individual-level observations, to increase representativeness of the sample. See text for details.

Table A28: Parent and Child Occupations, Normal School and Asylum Counties, with Propensity Score Weights

	(1)	(2)	(3)	(4)	(5)	(6)
	Unskilled	Craft/Oper	White Collar	Non Occ	Farmer	Diff from Parent
Panel A: Parent Occupations 1920, Parents of 1920 6-15 Year-Old Black Children						
Grew up in Normal School County	0.005	-0.006	-0.011	-0.006	0.015	
	(0.028)	(0.019)	(0.006)	(0.011)	(0.028)	
Observations	195	195	195	195	195	
R-Squared	0.270	0.356	0.226	0.099	0.658	
Mean DV, Asylum Counties	0.599	0.181	0.049	0.049	0.164	
p-value randomization inference	0.873	0.796	0.091	0.5	0.549	
Panel B: Child Occupations 1940, 1920 6-15 Year-Old Black Children						
Grew up in Normal School County	-0.024	-0.016	0.003	0.014	0.017^{+}	0.001
	(0.018)	(0.015)	(0.012)	(0.013)	(0.009)	(0.022)
Observations	195	195	195	195	195	195
R-Squared	0.235	0.150	0.245	0.131	0.581	0.194
Mean DV, Asylum Counties	0.577	0.205	0.119	0.113	0.042	0.542
p-value randomization inference	0.208	0.374	0.761	0.289	0.045	0.958

Notes: $^{+}$ $p < 0.1$, * $p < .05$, ** $p < .01$. This table is the equivalent of Table A12, but county-level averages are constructed from weighted individual-level observations, to increase representativeness of the sample. See text for details.

Table A29: Father and Son Occupations, Normal School and Asylum Counties, with Propensity Score Weights

	(1)	(2)	(3)	(4)	(5)	(6)
	Unskilled	Craft/Oper	White Collar	Non Occ	Farmer	Diff from Father
Panel A: Father Occupations 1920, Fathers of 1920 6-15 Year-Old Black Males						
Grew up in Normal School County	-0.010	0.011	-0.005	-0.011	0.024	
	(0.036)	(0.024)	(0.008)	(0.011)	(0.035)	
Observations	164	164	164	164	164	
R-Squared	0.318	0.339	0.207	0.124	0.615	
Mean DV, Asylum Counties	0.502	0.169	0.043	0.025	0.234	
p-value randomization inference	0.813	0.692	0.472	0.129	0.505	
Panel B: Son Occupations 1940, 1920 6-15 Year-Old Black Males						
Grew up in Normal School County	-0.014	-0.011	0.011	-0.009	0.024^{+}	-0.018
	(0.024)	(0.022)	(0.014)	(0.013)	(0.012)	(0.026)
Observations	164	164	164	164	164	164
R-Squared	0.124	0.121	0.252	0.112	0.491	0.205
Mean DV, Asylum Counties	0.531	0.229	0.100	0.085	0.056	0.585
p-value randomization inference	0.612	0.654	0.462	0.46	0.075	0.523

Notes: $^{+}$ $p < 0.1$, * $p < .05$, ** $p < .01$. This table is the equivalent of Table A13, but county-level averages are constructed from weighted individual-level observations, to increase representativeness of the sample. See text for details.

Table A30: 1940 Outcomes of 1920 6-15 Year-Old White Children, County-Level Outcomes based on 1920 County, with Propensity Score Weights

	(1)	(2)	(3)	(4)	(5)	(6)
	$\geq$ HS	$\geq$ Some College	$\geq$ College	Married	Employed	Ln(HH Income)
Panel A: Males, Lower Socioeconomic Status						
Grew up in Normal School County	0.007	0.007^{+}	0.001	0.005	-0.006^{+}	-0.026
	(0.007)	(0.004)	(0.002)	(0.004)	(0.003)	(0.016)
Observations	314	314	314	314	314	314
R-Squared	0.575	0.482	0.409	0.636	0.367	0.489
Mean DV, Asylum Counties	0.307	0.113	0.055	0.705	0.900	7.028
Panel B: Females, Lower Socioeconomic Status						
Grew up in Normal School County	0.008	0.023**	0.008**	0.020*	-0.022**	-0.012
	(0.010)	(0.006)	(0.003)	(0.008)	(0.007)	(0.015)
Observations	314	314	314	314	314	314
R-Squared	0.452	0.395	0.329	0.678	0.668	0.451
Mean DV, Asylum Counties	0.399	0.137	0.046	0.692	0.305	7.049
Panel C: Males, Higher Socioeconomic Status						
Grew up in Normal School County	0.016*	0.025**	0.013**	0.001	-0.003	0.001
	(0.006)	(0.005)	(0.004)	(0.005)	(0.003)	(0.012)
Observations	314	314	314	314	314	314
R-Squared	0.582	0.560	0.364	0.529	0.194	0.374
Mean DV, Asylum Counties	0.499	0.233	0.124	0.704	0.897	7.231
Panel D: Females, Higher Socioeconomic Status						
Grew up in Normal School County	0.021*	0.045**	0.024**	0.013^{+}	-0.015*	-0.004
	(0.008)	(0.008)	(0.005)	(0.008)	(0.007)	(0.010)
Observations	314	314	314	314	314	314
R-Squared	0.512	0.532	0.450	0.700	0.646	0.430
Mean DV, Asylum Counties	0.567	0.234	0.101	0.645	0.369	7.220

Notes: $^{+}$ $p < 0.1$, * $p < .05$, ** $p < .01$. This table is the equivalent of Table A15, but county-level averages are constructed from weighted individual-level observations, to increase representativeness of the sample. See text for details.

Table A31: 1940 Outcomes of 1920 6-15 Year-Old Black Children, County-Level Outcomes based on 1920 County, with Propensity Score Weights

	(1)	(2)	(3)	(4)	(5)	(6)
	$\geq$ HS	$\geq$ Some College	$\geq$ College	Married	Employed	Ln(HH Income)
Panel A: Males, Lower Socioeconomic Status						
Grew up in Normal School County	0.025	0.039*	0.002	0.031^{+}	0.014	0.064
	(0.019)	(0.018)	(0.004)	(0.018)	(0.016)	(0.051)
Observations	142	142	142	142	142	142
R-Squared	0.375	0.169	0.163	0.365	0.251	0.331
Mean DV, Asylum Counties	0.180	0.054	0.020	0.580	0.806	6.449
p-value randomization inference	0.27	0.058	0.524	0.101	0.368	0.203
Panel B: Females, Lower Socioeconomic Status						
Grew up in Normal School County	-0.016	0.018	0.003	0.031	-0.017	0.017
	(0.021)	(0.012)	(0.009)	(0.020)	(0.025)	(0.051)
Observations	77	77	77	77	77	77
R-Squared	0.417	0.384	0.255	0.513	0.216	0.423
Mean DV, Asylum Counties	0.197	0.081	0.039	0.312	0.523	6.037
p-value randomization inference	0.499	0.173	0.744	0.163	0.527	0.751
Panel C: Males, Higher Socioeconomic Status						
Grew up in Normal School County	0.024	0.041	0.036	-0.016	0.018	-0.142
	(0.038)	(0.030)	(0.030)	(0.035)	(0.027)	(0.087)
Observations	93	93	93	93	93	93
R-Squared	0.189	0.250	0.163	0.146	0.114	0.111
Mean DV, Asylum Counties	0.279	0.123	0.053	0.595	0.809	6.639
p-value randomization inference	0.603	0.255	0.325	0.698	0.567	0.242
Panel D: Females, Higher Socioeconomic Status						
Grew up in Normal School County	-0.003	-0.010	0.018	0.028	-0.031	-0.166
	(0.051)	(0.045)	(0.029)	(0.036)	(0.053)	(0.129)
Observations	50	50	50	50	50	49
R-Squared	0.231	0.292	0.342	0.252	0.173	0.190
Mean DV, Asylum Counties	0.358	0.181	0.077	0.220	0.565	6.364
p-value randomization inference	0.963	0.866	0.572	0.544	0.615	0.199

Notes: $^{+}$ $p < 0.1$, * $p < .05$, ** $p < .01$. This table is the equivalent of Table A16, but county-level averages are constructed from weighted individual-level observations, to increase representativeness of the sample. See text for details.

B.5 Alternative Historical Mobility Measure: Conditional Mean Occupational Income Score Rank

In this section, we supplement our analysis of historical economic mobility using an alternative measure. Similar to Tan (2023), we construct predicted occupational income score ranks of the children, for different moments of the father's occupational income score rank distribution. An important caveat here is that the fraction of farmers in 1850 is very high, which may make the results somewhat harder to interpret. Table A6 shows that roughly 71% of fathers in the asylum counties in our sample are farmers, and on average this is 5 percentage points higher in normal school counties.

We construct occupational income score ranks using the 1950 occupational income scores from IPUMS. Tan (2023) and Collins and Wanamaker (2022) construct occupation scores using 1940 incomes, so that they are based on incomes closer to their period in question. However, as described in those papers, there are no good measures of farmers' incomes in 1940 (as only wage and salary income is available in that year). As a result, in those papers the 1940 occupational income score for farmers is imputed based on the assumption of a constant ratio between 1940 and 1960 of total compensation for farmers to farm laborers. Given our interest in calculating parental ranks in 1850, this seems less reasonable and like it would make interpretation much more difficult. As a result, we use the 1950 occupational income scores, which have the benefit of better income measurement for farmers, but the cost of being 10 years farther from our year of interest.

Our preferred specifications in Tables 1 and 2 do not rely as heavily on occupational income scores. Table 1 does not use these at all, and Table 2 only to split the sample into high- and low-SES groups. Given the issues with using occupational income scores in our setting discussed above, we think this is an advantage of our main specifications.

Based on recent literature suggesting flexible occupation scores are important for cross-group comparisons, Tan (2023) and Collins and Wanamaker (2022) construct occupation scores based on average earnings of people in the same occupation, race, region of residence, and gender. These concerns are arguably less relevant for our setting, as comparison across racial groups, gender, and region is not our focus. In this historical analysis, our focus is on White males within a state, given the small sample sizes for Black males.

We restrict to our linked samples of sons and fathers. We follow Tan (2023) and for this exercise we drop sons and fathers with missing occupation or with occupations which have not yet been classified. We obtain the rank of the individual's occupation score among other individuals their age (or among other fathers with sons the same age), and then normalize the rank by the number of individuals of that age, yielding a percentile rank for sons and for fathers.

We follow Tan (2023) and estimate regressions separately for each county (Tan (2023) used commuting zones), in which the dependent variable is the son's rank, and the independent variable is the father's rank. We then obtain the predicted son's rank in each county for fathers who are at the 25th, 50th, and 75th percentiles. We then use these predicted ranks as dependent variables, and compare within states across normal school and asylum counties.

Tables A32 and A33 show there is no statistically significant difference in son's predicted occupational score rank in normal school counties versus asylum counties in the same state, for White sons with fathers' occupational score rank at the 25th, 50th, or 75th percentile, for children in 1850 or 1920. Table A34 shows that sons' predicted rank is lower in normal school counties by roughly 1 to 1.5 percentiles for Black children in 1920 with fathers at the 25th and 50th percentiles. Thus, with the important caveats mentioned above, these results are consistent with our results in Tables 1 and 2 showing little difference in mobility in 1850-1860 and little difference in economic mobility based on labor market outcomes in 1920-1940.

Table A32: Sons' predicted occupational score percentile in 1860 conditional on father's in 1850, White children

	(1)	(2)	(3)
Grew up in Normal School County	-0.009	-0.000	0.009
	(0.013)	(0.013)	(0.023)
Observations	102	102	102
R-squared	0.366	0.310	0.246
Mean DV, Asylum Counties	0.420	0.474	0.528
p-value randomization inference	0.52	0.998	0.748
Father's Occupation Score Percentile	25	50	75

Notes: $^{+}$ $p < 0.1$, * $p < .05$, ** $p < .01$. Robust standard errors in parentheses. See text for details.

Table A33: Sons' predicted occupational score percentile in 1940 conditional on father's in 1920, White children

	(1)	(2)	(3)
Grew up in Normal School County	-0.005	-0.003	0.000
	(0.004)	(0.003)	(0.003)
Observations	314	314	314
R-squared	0.618	0.519	0.332
Mean DV, Asylum Counties	0.405	0.475	0.545
Father's Occupation Score Percentile	25	50	75

Notes: $^{+}$ $p < 0.1$, * $p < .05$, ** $p < .01$. Standard errors clustered at the state level in parentheses. See text for details.

Table A34: Sons' predicted occupational score percentile in 1940 conditional on father's in 1920, Black children

	(1)	(2)	(3)
Grew up in Normal School County	-0.013**	-0.008*	-0.004
	(0.006)	(0.005)	(0.005)
Observations	164	164	164
R-squared	0.639	0.526	0.248
Mean DV, Asylum Counties	0.406	0.476	0.546
p-value randomization inference	0.03	0.088	0.459
Father's Occupation Score Percentile	25	50	75

Notes: $^{+}$ $p < 0.1$, * $p < .05$, ** $p < .01$. See text for details. Robust standard errors in parentheses.

C Differences in County Characteristics

We find that regional public universities affect education and mobility for children growing up in their local communities. The most natural explanation is that these universities reduce geographic frictions in college attendance, and this affects college attainment as well as income mobility and social outcomes. However, regional public universities may impact these outcomes through other channels as well. For example, regional universities may impact local economic outcomes, such as industrial composition, in ways that increase the return to education in the local community, and this may increase high school and college attainment. Regional universities may also affect other characteristics of the local community, such as the income distribution or family composition, which may affect mobility directly or indirectly, for example through affecting primary and secondary school quality.

C.1 Differences in County Characteristics in 1940

As an additional way of testing whether the previous normal schools were having an impact on their local economies before the birth cohorts in the Opportunity Insights data, we examine 1940 county characteristics. As we discussed, most of these institutions were colleges in 1940 offering at least a BA in education. Some were offering degrees in addition to education, and had been converted to a regional college (Figure A1a). We use the 1940 full count census data (Ruggles et al., 2021*a*), as well as county-level data from NHGIS (Manson et al., 2023).

Tables A35 and A36 show there are few statistically significant differences between normal school and asylum counties in population, dwelling characteristics, labor market statistics, occupation, or industry. The fraction employed in manufacturing in normal school counties was 2.4 percentage points lower, or roughly 12%, though the fraction employed in mining was .9 percentage points higher in normal school counties (Table A36). The fraction employed in agriculture, forestry and fishing in normal school counties was higher by 2.8 percentage points, or roughly 13%, but this is not statistically significant.

Table A35 shows results using as a dependent variable the first principal component of all the variables in the table. This component captures roughly 48% of the total variance. All variables have positive loadings except the fraction of the total population that is Black, male, and employed (14 and over), though for this last variable the loading is very close to zero. Other than these variables with negative loadings (that are also small), the magnitudes of the loadings are quite similar for all variables.

Table A36 shows results using as a dependent variable the first principal component of all the variables in the table. This component captures roughly 38% of the total variance. All variables enter with positive loadings, except fraction farmer and farm-manager, fraction employed in the agriculture, forestry, or fishing industry, fraction employed in mining, and fraction with occupation in domestic service (though the last two have very small loadings). The following variables have the highest loadings (around .3 in absolute value): fraction with semiprofessional occupation, fraction farmer or farm-manager, fraction clerical, sales, and kindred occupations, fraction employed in wholesale and retail trade, fraction employed in agriculture, and fraction employed in finance, insurance, and real estate.

While the overall economy is similar in normal school and asylum counties in 1940, there

Table A35: Population, Labor Force, and Dwelling Characteristics in 1940: Normal School and Asylum Counties

	Normal	Asylum	Within-State Difference
Population	121743.01	174916.86	-38,335.178
	[263443.58]	[474510.93]	(43,803.374)
Fraction of Total Population			
Urban	0.45	0.51	-0.030
	[.27]	[.23]	(0.033)
Living in Cities 25,000 and Over	0.19	0.27	-0.058
	[.3]	[.33]	(0.042)
Male	0.5	0.51	-0.001
	[.02]	[.02]	(0.003)
Black	0.09	0.07	-0.000
	[.16]	[.12]	(0.007)
Foreign	0.06	0.07	-0.011*
	[.06]	[.06]	(0.005)
Ln(Average Wage, Manufacturing)	6.83	6.87	0.019
	[.29]	[.28]	(0.031)
Ln(Value Added, Manufacturing)	15.22	15.5	-0.241
	[2.22]	[2.22]	(0.266)
Ln(Manufacturing Establishments)	3.93	4.24	-0.297
	[1.56]	[1.53]	(0.198)
Fraction Employed, Age 14 and Over	0.86	0.86	-0.004
	[.05]	[.05]	(0.006)
Fraction Seeking Work, Age 14 and Over	0.09	0.08	-0.001
	[.03]	[.04]	(0.004)
Fraction of Occupied Dwelling Units Occupied by Owners	0.48	0.49	0.002
	[.11]	[.1]	(0.011)
Fraction of Occupied Dwelling Units with			
Electric Lighting Equipment	0.71	0.77	-0.022
	[.23]	[.19]	(0.017)
Radio	0.79	0.84	-0.012
	[.18]	[.14]	(0.009)
Mechanical Refrigeration Equipment	0.36	0.41	-0.020
	[.16]	[.15]	(0.014)
Ln(Average Value of Owner Occupied Housing)	7.9	7.98	-0.050
	[.4]	[.42]	(0.042)
Ln(Median Value of Owner Occupied Housing)	7.65	7.79	-0.076
	[.53]	[.48]	(0.047)
Ln(Value of Crops Harvested)	14.82	14.87	0.013
	[.96]	[.81]	(0.107)
First Principal Component	-0.18	0.4	-0.325
	[2.94]	[2.72]	(0.261)

Notes: Data are county-level data from NHGIS (Manson et al., 2023). Standard errors clustered at the state level. The first principal component in the last row is the first principal component of all the variables in this table. This component captures roughly 48% of the total variance. All variables have positive loadings except the fraction of the total population that is Black, male, and employed (14 and over), though for this last variable the loading is very close to zero. Other than these variables with negative loadings (that are also small), the magnitudes of the loadings are quite similar for most variables.

Table A36: Occupation and Industry in 1940: Normal School and Asylum Counties

	Normal	Asylum	Within-State Difference
Fraction of Employment by Occupation			
Professional	0.07	0.07	-0.0004
	[.01]	[.02]	(0.002)
Semiprofessional	0.01	0.01	-0.0003
	[.004]	[.004]	(0.001)
Farmers and farm-managers	0.16	0.14	0.025^{+}
	[.13]	[.12]	(0.013)
Proprietors, managers, and officials (except farm)	0.08	0.08	0.002
	[.02]	[.02]	(0.002)
Clerical, sales, and kindred workers	0.13	0.14	-0.005
	[.06]	[.06]	(0.007)
Craftsmen, foremen, and kindred workers	0.1	0.11	-0.004
	[.04]	[.04]	(0.004)
Operatives and kindred workers	0.16	0.16	-0.003
	[.09]	[.08]	(0.008)
Domestic service	0.05	0.05	-0.002
	[.03]	[.02]	(0.002)
Service, except domestic	0.07	0.08	-0.016**
	[.03]	[.03]	(0.003)
Laborer (except farm)	0.07	0.07	-0.005
	[.04]	[.04]	(0.004)
Fraction of Employment by Industry			
Agriculture, Forestry, and Fishing	0.25	0.22	0.028
	[.17]	[.17]	(0.020)
Mining	0.02	0.01	0.009*
	[.05]	[.03]	(0.004)
Construction	0.09	0.09	0.001
	[.03]	[.02]	(0.003)
Manufacturing	0.18	0.2	-0.024*
	[.11]	[.12]	(0.010)
Transportation, Communication, and Other Utilities	0.06	0.06	0.002
	[.03]	[.03]	(0.003)
Wholesale and Retail Trade	0.14	0.15	-0.001
	[.04]	[.03]	(0.005)
Finance, Insurance, and Real Estate	0.02	0.02	-0.001
	[.01]	[.01]	(0.002)
Business and Repair Services	0.02	0.02	0.0004
	[.004]	[.003]	-0.0004
Personal Services	0.08	0.08	-0.003
	[.03]	[.03]	(0.003)
Professional and Related Services	0.08	0.08	-0.007**
	[.02]	[.03]	(0.002)
Public Administration	0.03	0.04	-0.004
	[.03]	[.04]	(0.003)
First Principal Component	-0.29	0.47	-0.389
	[2.88]	[2.67]	(0.345)

Notes: Occupation data are based on county-level data from NHGIS (Manson et al., 2023), and industry data are based on employed people in the 1940 full count census data (Ruggles et al., 2021*a*). Entertainment and Recreation Services is not shown in the table. The fraction employed is 1% in both normal school and asylum counties. Standard errors clustered at the state level. See text for details.

Table A37: School Attendance and Educational Attainment in 1940: Normal School and Asylum Counties

	Normal	Asylum	Within-State Difference
Fraction Attending School by Age			
5 to 6	0.42	0.44	0.001
	[.16]	[.15]	(0.008)
7 to 13	0.96	0.96	-0.001
	[.07]	[.04]	(0.004)
14 to 15	0.9	0.91	0.004
	[.07]	[.06]	(0.005)
16 to 17	0.69	0.7	0.016*
	[.11]	[.1]	(0.007)
18 to 20	0.26	0.25	0.025**
	[.07]	[.06]	(0.006)
21 to 24	0.06	0.06	0.007^{+}
	[.03]	[.03]	(0.004)
Fraction of 16 to 17 year-olds			
Attending school, completed HS	0.05	0.04	0.007*
	[.02]	[.02]	(0.003)
Attending school, did not yet complete HS	0.63	0.65	0.010
	[.11]	[.1]	(0.007)
Median School Years Completed, Age 25 and Over			
Male	8.09	8.27	0.041
	[.93]	[.72]	(0.064)
Female	8.65	8.75	0.160*
	[1.08]	[.88]	(0.076)
Fraction Age 25 and Over with			
Four or more years of college	0.05	0.04	0.005*
	[.01]	[.02]	(0.002)
One to three years of college	0.15	0.15	0.009**
	[.04]	[.04]	(0.003)
Four years of high school	0.13	0.14	0.001
	[.04]	[.04]	(0.004)

Notes: Data are county-level data from NHGIS (Manson et al., 2023), except the decomposition of the fraction of 16-17 year-old individuals attending school, which is based on the 1940 full-count census data (Ruggles et al., 2021*a*). Standard errors clustered at the state level. See text for details.

are differences in education (Table A37). School attendance rates are similar for children through age 15. However, attendance rates are higher for individuals age 16-17, 18-20, and 21-24. For these latter two age groups, this may reflect that the respondents are college students who moved to the county to attend college. For 16-17 year old individuals, this may reflect greater high school attendance, or that some people in this age group are attending college at a young age and moved from another county. To test this, we use the 1940 full count census data (Ruggles et al., 2021*a*), and decompose the 16-17 year-old school attendance into individuals who completed high school (using the educational attainment variable) and separately those who had not completed high school. We keep the denominator equal to the total number of 16-17 year-old individuals in both variables. We see evidence that the greater attendance rate of 16-17 year-old individuals is partly explained (with a coefficient of .007) by students who have already completed high school (and may have moved from other counties), and also partly explained (with a coefficient of .01) by students who had not yet completed high school and are more likely to be living in the county with their family. This suggests high school attendance rates are higher for children growing up in normal school counties in 1940, though the difference is not large (1 percentage point higher from a mean of roughly 65 percent, or 1.5%). This is consistent with our evidence on high school graduation rates using the linked 1920-1940 censuses in Section C.1.

We also see that while there is no difference in the fraction of individuals 25 and over with exactly four years of high school, a greater share of individuals in normal school counties have one to three years of college (.9 percentage points higher or 6%) and four or more years of college (.5 percentage points higher or roughly 12.5%). In 1980, the difference was 2 percentage points, but roughly the same relative effect of 13% given the higher baseline mean in 1980.

Together, this evidence suggests that the institutions that had been the normal schools were not having dramatic impacts on their counties in 1940, with the exception of modest increases in educational attainment.

C.2 Differences in 8th and 9th Grade Attainment for Children Living with Low-Educational-Attainment Parents, 1940

In the table below, we show differences in educational mobility between normal school and asylum counties in 1940, using data from Card, Domnisoru and Taylor (2022). Panel A shows results when the outcome is the county-level fraction of children attaining eighth grade, living with parents with grade six maximal educational attainment. Panel B shows results when the outcome is the county-level fraction of children attaining ninth grade, living with parents with grade six maximal educational attainment. Table A38 shows there are no significant differences in upward educational mobility in 1940 using these measures. While it is possible the normal schools would have an impact on 8th and 9th grade attainment through raising the quality of the K-12 education system, the most direct effects would seem to be on older ages. Indeed, we do see effects on college attainment for children in 1920 when they are observed in 1940 (Tables 2 and A15). Finding insignificant differences in 8th and 9th grade attainment for children in households with low educational attainment is also consistent with lack of sorting of high economic mobility households into normal school counties relative to asylum counties.

Table A38: Measures of Upward Educational Mobility, 1940

	Normal School	Asylum	Within-State Difference
Fraction attaining 8th grade, living with parents with grade six maximal education			
White	0.7	0.75	0.00
	[.18]	[.16]	(0.01)
Black	0.58	0.66	0.02
	[.29]	[.26]	(0.03)
Fraction attaining 9th grade, living with parents with grade six maximal education			
White	0.49	0.52	0.01
	[.16]	[.15]	(0.01)
Black	0.44	0.5	0.00
	[.28]	[.25]	(0.03)

Notes: Columns 1 and 2 show the mean and standard deviation for normal school and asylum counties. Column 3 shows the coefficient on normal school county, when the dependent variable is the county mobility measure, and we include state fixed effects. These mobility measures are from Card, Domnisoru and Taylor (2022). We show standard errors clustered at the state level in parentheses in column 3. For White individuals, there are 203 normal school counties and 121 asylum counties in column 3. For Black individuals, there are 137 normal school counties and 79 asylum counties in column 3.
$^{+}$ $p < 0.1$, * $p < .05$, ** $p < .01$

C.3 Differential characteristics of normal school counties using data from Chetty et al. (2018) and Chetty and Hendren (2018)

In this section, we test for differences between normal school and asylum counties in more recent years, on a number of characteristics related to the mechanisms discussed at the beginning of section C. We use data from Chetty et al. (2018) and Chetty and Hendren (2018).[56]

Consistent with Howard, Weinstein and Yang (2024) we find very little within-state difference in economic characteristics between normal school and asylum counties (Table A39). We focus on economic characteristics, K-12 public schools and colleges, and family characteristics. For economic characteristics, in Howard, Weinstein and Yang (2024) we show that in normal school counties the employment share in accommodations and food services is about 1 percentage point higher (significant at the 1 percent level), the share in retail trade is higher by about 0.6 percentage points, and the share in wholesale trade and finance and insurance are both lower by about 0.4 percentage points. These differences are small, and none of them suggest jobs with a higher return to college degrees in normal school counties. In this paper, we show in Table A39 that there is no difference in wage growth for high school graduates, or overall job growth. As we also show in Howard, Weinstein and Yang (2024) we see higher bachelor's degree share by about 3 percentage points in normal school counties. Higher bachelor's share may affect education levels of lower-income children in several ways, one of which is the quality of the local public elementary and secondary schools.

For educational characteristics, there is no difference in expenditures per student, or in 3rd grade math scores, however, the student-teacher ratio is modestly lower in normal school counties by about 0.5 (approximately 3%). This may suggest other differences in local schools that affect high school graduation and college enrollment rates in normal school relative to asylum counties. Consistent with regional public universities affecting outcomes by making a local college education more affordable, the tuition at colleges in the county is lower by about $2500 in normal school counties, which is roughly 36 percent lower. Section A.1 shows many more differences between universities in normal school and asylum counties.

Children living in low-income households in normal school counties are more likely to have two parents whose income together is the same as single parents' income in asylum counties. The fraction of children claimed by two people as a dependent, among those whose parents are at the 25th income percentile, is higher by 3 percentage points in normal school counties, which is roughly 6 percent higher based on the average in asylum counties.[57] There is no difference for children whose parents are at the 75th percentile. As regional public universities raise education levels and marriage of children from lower-income families, they may also have done so for their parents. In this case, some of the effect of regional public

[56] We focus on variables that do not come from the census, given that students are included in the census in the location where they live as students, and this will affect per capita estimates.

[57] This does not say that children of low-income parents are more likely to live with both parents in normal school counties than asylum counties because this statement is dependent on the total income of their parents being at the 25th percentile, which is endogenous to the number of people claiming the child as a dependent on their tax forms.

Table A39: Chetty et al. (2018) and Chetty and Hendren (2018) Covariates

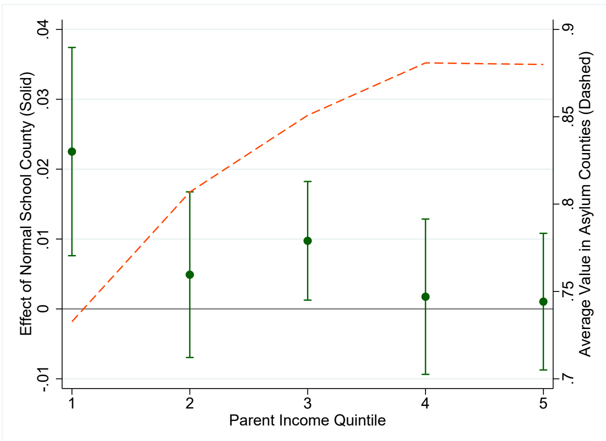
	Normal	Asylum	Within-State Difference
Economic Characteristics			
Manufacturing employment share, 2000	0.13	0.15	-0.01
	[.06]	[.07]	(0.01)
Average annualized job growth, 2004-2013	0	0	0.00
	[.01]	[.01]	(0.00)
HS grad. wage growth, 2005-2009 - 2010-2014	0.06	0.05	0.01
	[.11]	[.07]	(0.01)
Bachelor's degree share, age $\geq$ 25, 2000	0.24	0.22	0.03*
	[.07]	[.09]	(0.01)
Population, 2000	269,614	305,908	-30,670.68
	[765,738]	[593,196]	(92,294)
Children < 18, 2000	67,974	77,295	-8,224.44
	[209,691]	[152,889]	(24,235)
K-12 Public Schools and Colleges			
K-12 expenditures per stud., 1996-1997	6.38	6.38	-0.00
	[1.43]	[1.41]	(0.07)
K-12 student teacher ratio, 1996-1997	16.88	17.52	-0.46*
	[2.18]	[2.13]	(0.18)
Mean 3rd grade math test scores, 2013	3.28	3.29	0.03
	[.63]	[.71]	(0.07)
College tuition, local colleges, IPEDS 2000	4149.01	6778.76	-2,455.74**
	[3,836.2]	[4,664.33]	(589.25)
Family characteristics, children in Chetty et al. (2018)			
Children claimed by two people			
parent income at p25	0.51	0.49	0.03*
	[.12]	[.12]	(0.01)
parent income at p75	0.94	0.93	0.00
	[.04]	[.06]	(0.01)
Fraction of childhood spent in the county	0.74	0.76	-0.01^{+}
	[.07]	[.06]	(0.01)

Notes: Columns 1 and 2 show mean and standard deviation of county characteristics for normal school and asylum counties. Column 3 shows the coefficient on normal school county, when the dependent variable is the county characteristic, and we include state fixed effects. We show standard errors clustered at the state level in parentheses in column 3. All economic variables except county population are from Chetty et al. (2018). Variables related to K-12 public schools and colleges are from Chetty and Hendren (2018), except 3rd grade math scores which are from Chetty et al. (2018). Fraction of children claimed by two people as a dependent is from Chetty et al. (2018), and is based on parents of children in the 1978-1983 birth cohorts, and parents' average household adjusted gross income in 1994, 1995, and 1998-2000. Fraction of childhood spent in the county is from Chetty et al. (2018). $^{+}$ $p < 0.1$, * $p < .05$, ** $p < .01$.

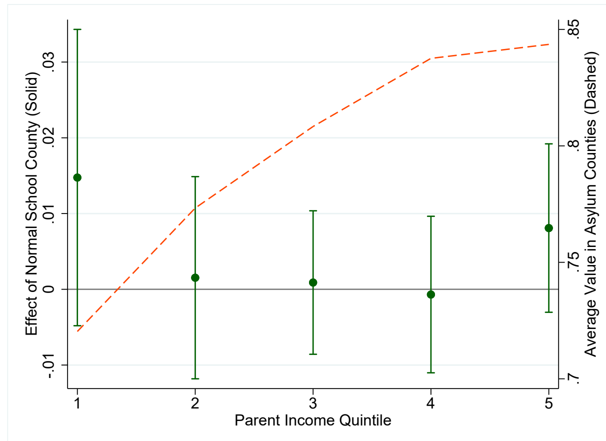
universities on children may come through the effect they had on the previous generation.[58]

Using data from Chetty et al. (2016), we provide suggestive evidence that the mobility effects are not driven by differences in likelihood of having two parents. These data are similar to the other outcome data we use, but further disaggregate outcomes by whether children have one or two parents who claim them on their taxes. The only outcomes available are regarding the likelihood of employment, and only disaggregated by gender. Among those who have two parents claim them on their taxes, we show normal school assignment increases employment by two percentage points for men whose parents are in the first income quintile (Appendix Figure A3). The magnitude is similar for men with single parents, and 1.5 percentage points for women with two parents, though neither are statistically significant. These results suggest our main effects are not driven by differences between normal school and asylum counties in likelihood of having two parents during childhood.

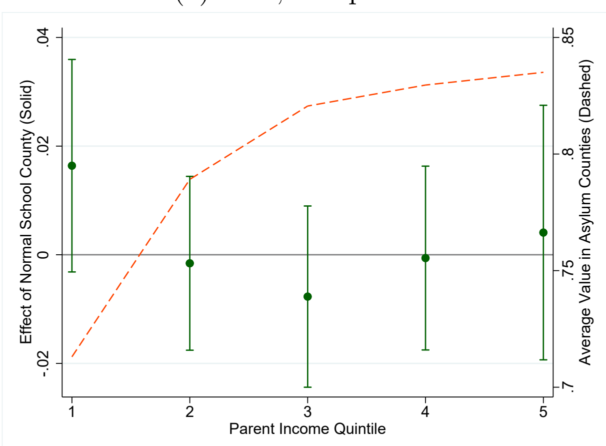
[58] We also find insignificant differences in racial and income segregation indices from Chetty and Hendren (2018).



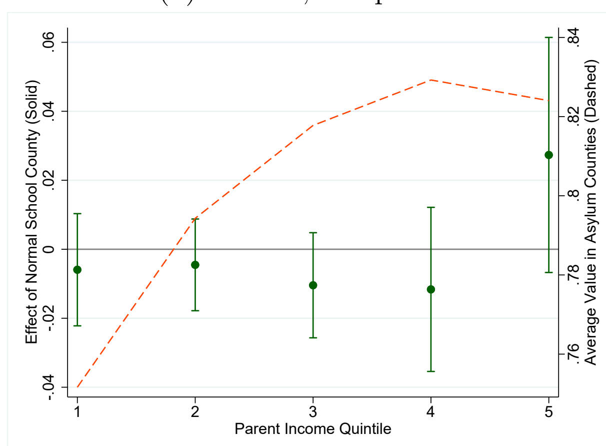
(a) Men, two parents



(b) Women, two parents



(c) Men, single parent



(d) Women, single parent

Figure A3: **Effect of a normal school on local children's age-30 employment for 1980-1982 birth cohorts, by sex and parental structure.** Dots with spikes represent the estimated effect of a normal school on the outcome variable, conditional on state fixed effects. Outcome variables are measured at different percentiles of the parents' income distribution, which is the x-axis. The spikes span the 95 percent confidence intervals. The estimates and confidence intervals correspond to the y-axis on the left-hand side of the figure. Standard errors are clustered at the state level. The dashed lines show the mean of the outcome variable in asylum counties, and correspond to the y-axis on the right-hand side of the figure.

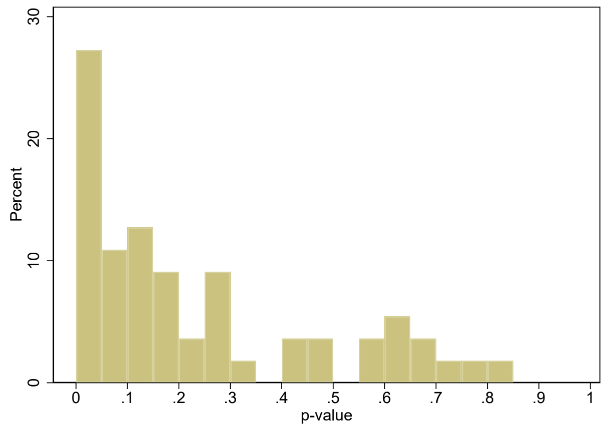


Figure A4: p-values from the 55 regressions. Each bin has a width of 0.05, so if the p-values were distributed uniformly, 5 percent would be in each bin.

D Additional Results using the Chetty et al. (2018) and Chetty and Hendren (2018) data

D.1 Additional Detail on Variables

The measure of teen motherhood is constructed based on whether a woman ever claims a dependent who was born while she was 13 to 19 years old. As Chetty et al. (2018) discuss, this is an imperfect measure since it relies on the woman claiming the child as a dependent at some point, but they document that this is aligned with estimates from the ACS. Staying with parents is defined as having an address that matches their parents' in 2015. Staying in childhood commuting zone is defined as the most recent commuting zone matching any commuting zone they lived in before 23.

D.2 Discussion of Multiple Hypothesis Testing

In our analysis of the Opportunity Insights data, we used our empirical strategy to investigate the effect of universities on 11 different outcomes at 5 different points of the parents' income distribution. A reader may reasonably wonder which takeaways are robust to considering multiple hypothesis testing. To give a general idea of the overall significance of our results, Figure A4 shows a distribution of the p-values for the 55 outcomes. More than a quarter of the unadjusted p-values are less than 5 percent, and another tenth are less than 10 percent. Of course, this is not a formal test, but is suggestive that universities have some causal effect.

To formally show that universities matter, we implement Young (2020), a randomization-based omnibus test to see if we can reject the null hypothesis that the normal schools have no effect on any of the 55 outcomes. The p-value of the randomization-c test is 0.0445. So the null hypothesis of no effect of the normal schools is rejected at conventional levels.[59]

[59] The randomization-t test has a p-value of .1044, but randomization-c is generally viewed as more appropriate when outcomes are highly dependent, as we would expect them to be here.

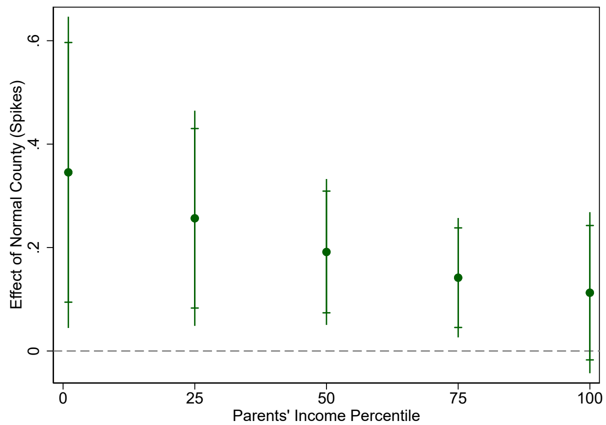


Figure A5: Principal Components

Given that there is *some* effect, we turn our focus to what the effect is. Before doing any econometrics, we must ask what makes this study interesting. The answer is not that regional universities affect any one particular outcome that we tested above. In our opinion, the main point of this paper is that universities affect "social mobility," i.e. they affect the common part of all the outcomes, and that they do so for children with low-income parents.

Based on wanting to test "social mobility," we create a measure that is the principal component of the 11 outcomes we have previously considered: having a college degree, attending college, having a high school degree, working, the percentile of family income, the percentile of individual income, marriage, teen birth, incarceration, living at home, and living outside of their childhood commuting zone.[60] We calculate this principal component treating each county in our sample by each percentile we consider as one observation. We then see if there is an effect on this principal component at each of the five percentiles. We adjust for the fact that this is five different tests by applying the Romano and Wolf (2005) correction for adjusting p-values.

The results of this procedure are in Figure A5.[61] The confidence intervals in the figure are not adjusted for multiple hypothesis testing. However, the p-values associated with each percentile, from the Romano and Wolf (2005) procedure, are for the 1st percentile, 0.041; for the 25th percentile, 0.040; for the 50th percentile, 0.024; for the 75th percentile, 0.040; and for the 100th percentile, 0.161. So even adjusting for multiple hypothesis testing, there is statistically significant evidence that universities have an effect on the principal component of these outcomes for all but the very top of the parents' income distribution.

The point estimates are bigger for children of lower-income parents, but we do not view the differences as a key aspect of our study. Whether or not universities help the outcomes of

[60]Our use of a principal component is distinct from Anderson (2008), who emphasized creating an index that overweights outcomes that are less correlated to the others. We are not interested in maximizing the power of our test, but think there is economic significance in the underlying factor that can explain the most variation across these eleven outcomes.

[61]The principal component has similar scoring coefficient magnitudes, between 0.22 and 0.34 for all 11 outcomes. Teen birth, incarceration, staying within the commuting zone, and staying at the parent's home have negative coefficients.

high-parental-income children, the fact that they help the outcomes of low-parental-income children implies that they improve social mobility for children that grow up near them, relative to the *national* distribution. As we discussed in the introduction, this is more of the policy purpose of the regional university, rather than whether they move up in the *local* distribution.

D.3 Comparing Close and Far Asylum Counties

Given that our evidence suggests asylum county children are less served by regional public universities because of geographic frictions, we might wonder if asylum counties that are nearby to normal school counties might experience more of the benefits of the normal schools. Are there spillovers from the normal school counties that are larger for close asylum counties compared to far asylum counties?

To do this analysis, for each asylum county, we calculated the distance to the closest normal school county. We then split the sample of asylum counties into two, based on the median distance to the closest normal school county. The median distance was 41 miles.

A map of the close and far asylums can be seen here. In this figure, red counties had normal schools, while yellow and green counties had asylums. Yellow counties are within 41 miles of normal school counties, while green counties are more than 41 miles.

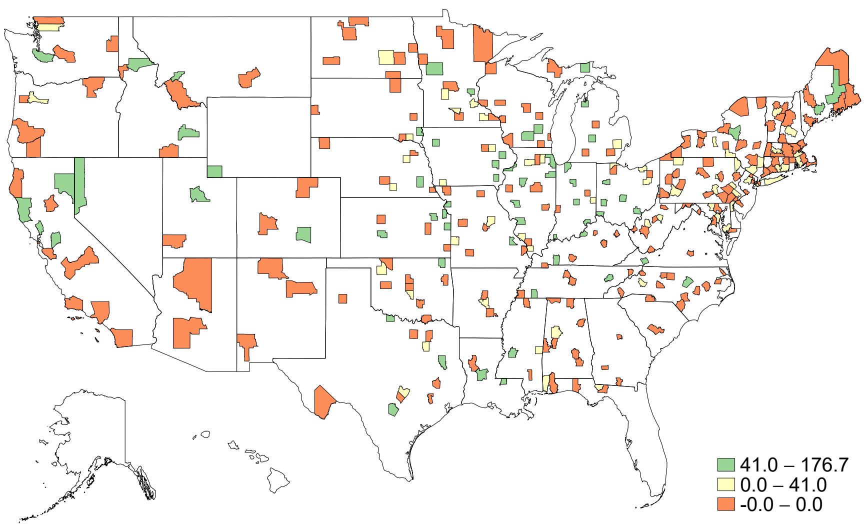


Figure A6: Map of normal schools, close asylums, and far asylums. Red counties have normal schools, yellow counties have close asylums, and green counties have far asylums. The legend denotes the number of miles between the county and the nearest normal county.

We then estimate two regressions

$$y_i = \beta_{\text{normal}} \text{Normal}_i + \alpha_s + \epsilon_i$$

where the sample consists of normal school and far asylum counties, and

$$y_i = \beta_{\text{close asylum}} \text{Close Asylum}_i + \alpha_s + \epsilon_i$$

where the sample consists of close and far asylum counties that are not also normal school counties. β_{normal} compares the outcome y between normal school counties and same-state far asylum counties, and $\beta_{\text{close asylum}}$ compares the outcome y between close asylum counties and same-state far asylum counties. We cluster standard errors at the state level. We estimate separate regressions to avoid the contamination bias discussed in Goldsmith-Pinkham, Hull and Kolesár (2024).[62]

First, we present the results where y is the first principal component, corresponding to Figure A5.

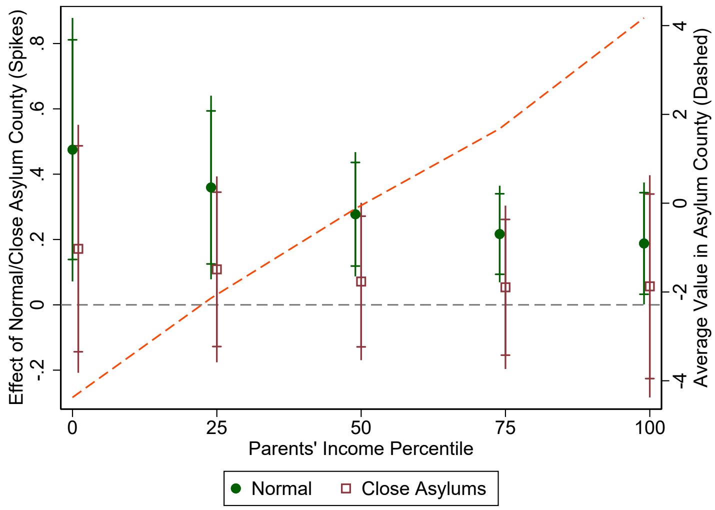


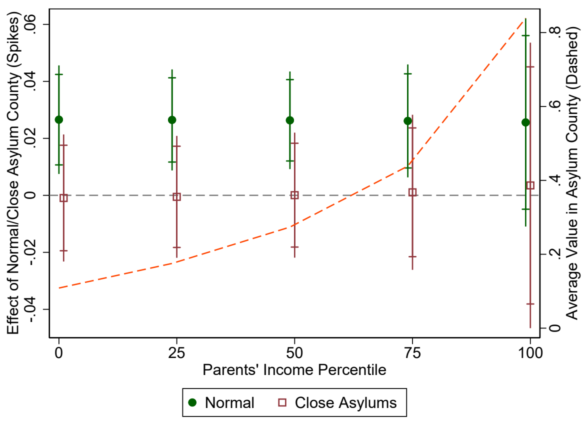
Figure A7: Comparison of Normal School vs. Far Asylums and Close vs. Far Asylums: Principal Component of Social Mobility

Figure A7 shows higher levels of social mobility in normal school counties relative to same-state far asylum counties. This can be seen by looking at the estimates in green, which are just a bit bigger—but not statistically distinguishable—from the estimates comparing normal school counties to all same-state asylum counties. The estimates marked with squares reflect the difference between the close and far asylum counties within the same state. If there are positive spillovers from the normal schools to the close asylum counties, these estimates will be positive. If the square is at zero, this suggests the effect of normal schools on close asylums is the same as the effect of normal schools on far asylums. If the square is at the same magnitude as the circle, this suggests spillovers from the normal school counties to the close asylums that are so large that the outcomes in the two types of counties are the same. It is evident that the red confidence intervals (for the squares) span both zero and the green point estimates.

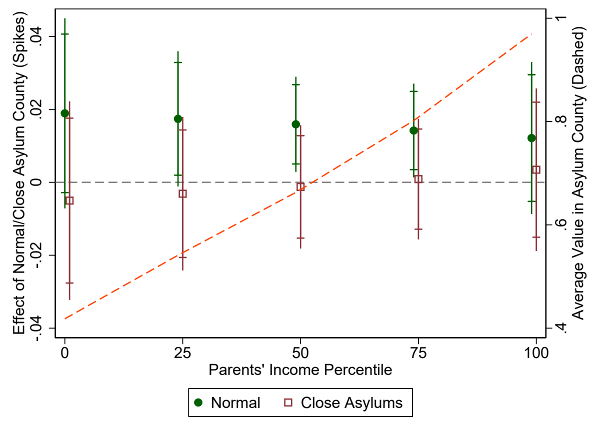
Thus, we cannot say very much about the effects on close asylums due to the large standard errors. We cannot rule out substantial spillovers from the normal school counties to the close asylum counties, including up to the same size as the effect of historical normal school assignment on the normal school counties themselves. However, we also cannot rule out zero spillovers.

[62]The two regressions we estimate in this section are also estimated on different sets of states relative to each other, and relative to the full sample in the main text. This is because 13 states have only close asylum counties, 12 states have only far asylum counties, and 15 states have normal schools, far asylums, and close asylums. Although we do not show it, we also estimate the same regressions limiting our sample to only the 15 states that have normal schools, far asylums, and close asylums. This addresses concerns of treatment effect heterogeneity by state, where that heterogeneity is correlated with the number of close or far asylums. The results are similar.

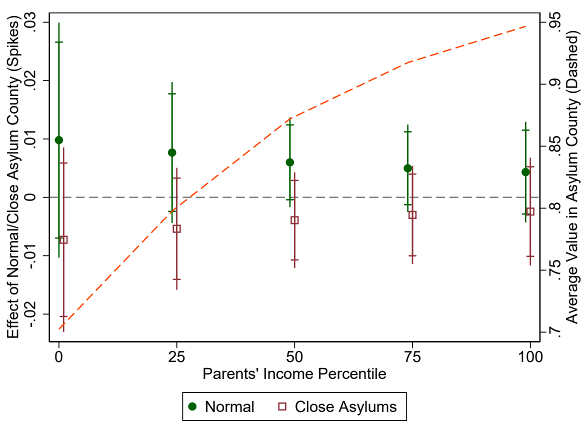
In Figures A8 and A9, we show all eleven variables that we look at as outcomes in the main paper. In the majority of cases we cannot rule out spillovers that are the size of the effect on normal school counties *or* zero.



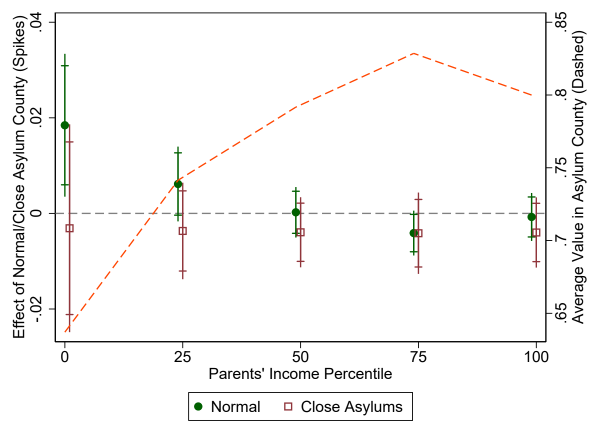
(a) College



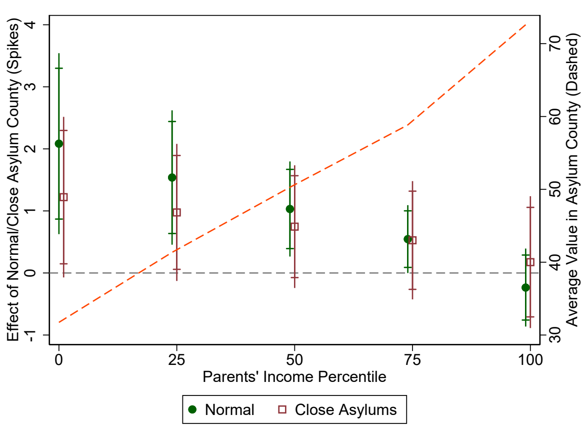
(b) Some College



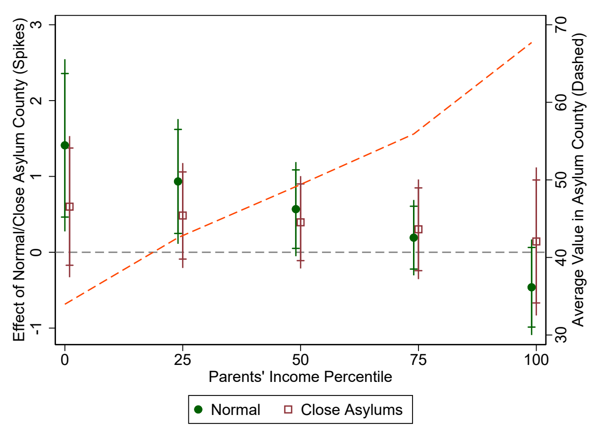
(c) High School



(d) Positive W-2 Earnings

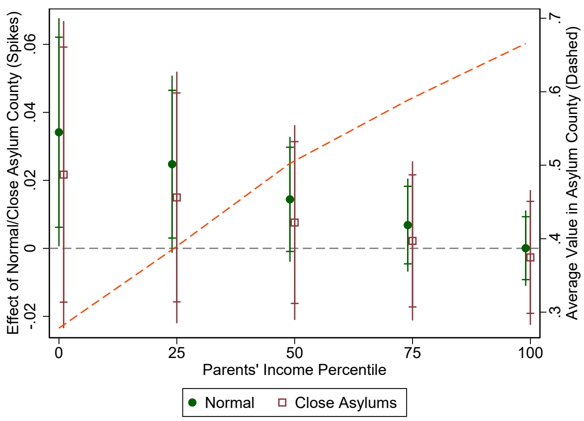


(e) Family Income Percentile

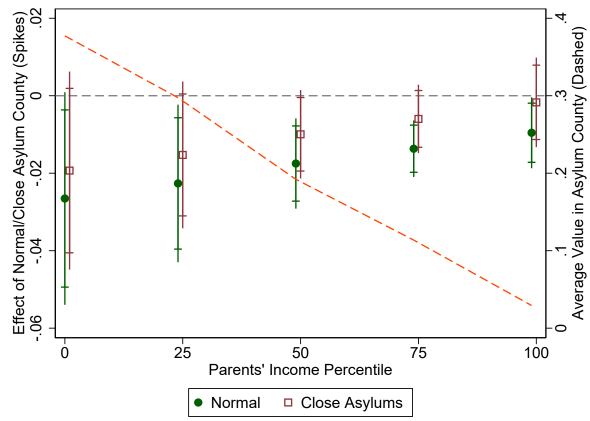


(f) Individual Income Percentile

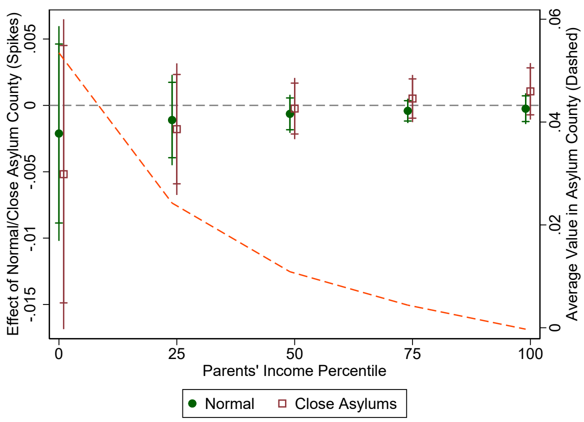
Figure A8: Comparison of Normal School vs. Far Asylums and Close vs. Far Asylums



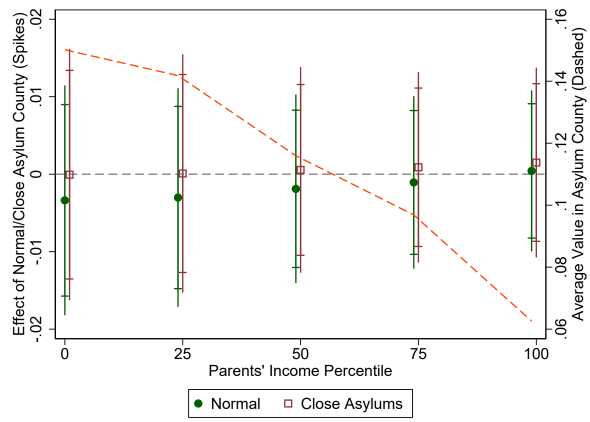
(a) Married



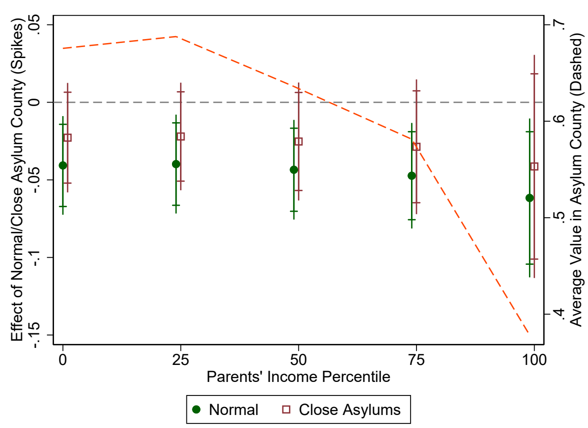
(b) Teen birth (women-only)



(c) Incarcerated



(d) Staying with Parents



(e) Live in Childhood Commuting Zone

Figure A9: Comparison of Normal School vs. Far Asylums and Close vs. Far Asylums

D.4 Detail on Comparing Chetty et al. (2018) and Chetty and Hendren (2018) Results

In Table A40, we show alternative specifications for Table 3, columns (1) and (2). In the main text, we used separate weighting schemes for the baseline results using the Chetty et al. (2018) data, in which the regressions were unweighted, and the causal-effects-on-people results using the Chetty and Hendren (2018) data, in which the results were weighted with precision weights. These are reproduced in column (1) and column (5) in Table A40.

In this section, we additionally show three alternative specifications. Column (2) uses the same data as Column (1) but the weights from Column (5). The point estimate is slightly lower, but the standard error increases by a factor of 2. The very large increase in standard error is why we do not prefer this regression for our main specification. Column (4) uses the same data as Column (5), but is unweighted as in Column (1). While columns (1) and (4) are both unweighted, a few observations have an outcome in the Chetty et al. (2018) data (column 1) but not the Chetty and Hendren (2018) data (column 4), so implicitly those counties get zero weight in column (4). Here, the point estimate also falls slightly, but the standard errors also increase. Chetty and Hendren (2018) suggests that the weights are necessary to account for the fact that some of the coefficients are quite noisy, so we prefer Column (5) as our main specification. Finally, as another check on the comparability of the two datasets, we also look at college attendance as measured in Chetty and Hendren (2018), but using the sample of permanent residents. Using this sample the effect can be interpreted as an effect on the place, and is measured per childhood, not per year. The differences between Columns (1) and (3) are how college attendance is measured, and also that Column (1) included people that lived in the county for part of their childhood, weighted to reflect how many years they spent there. Column (3) is a bit noisier, but the point estimate is actually larger, and still statistically significant. Overall, this exercise justifies why we prefer Columns (1) and (5): because they maximize power, but also shows that the positive point-estimates seem to be robust to alternative specifications.

Table A40: Effect on College Attendance, 25th percentile parental income, Robustness

	(1) Some College Age 25+	(2) Some College Age 25+	(3) Attended College Age 18-23	(4) Attended College Age 18-23	(5) Attended College Age 18-23
Normal	1.397^{*}	0.804	1.870^{*}	0.0725	0.143^{+}
	(0.680)	(1.204)	(0.864)	(0.0898)	(0.0764)
Observations	324	305	324	305	305
Weights	Unweighted	Precision Weights	Unweighted	Unweighted	Precision Weights
Scale	Per Childhood	Per Childhood	Per Childhood	Per Year	Per Year
Interpretation	Effect on Place	Effect on Place	Effect on Place	Effect on Person	Effect on Person

Standard errors clustered by state. $^{+}$ $p < 0.1$, * $p < .05$, ** $p < .01$. Outcome data in columns 1 and 2 are from Chetty et al. (2018), and outcome data in columns 3-5 are from Chetty and Hendren (2018).

Table A41 shows a similar analysis for income, and is an expanded version of Table 3 columns (3) and (4). Again, the alternative weighting schemes which make the regressions in Columns (1) and (5) more comparable are also less powerful, and while the point-estimates are still positive, they are not statistically significant. In this case, the point estimate on permanent residents in Column (3) does differ from our main results, but we cannot rule out a positive effect.

Table A41: Effect on Income Percentile, 25th percentile parental income, Robustness

	(1) Family Income Percentile, 2014-15	(2) Family Income Percentile, 2014-15	(3) Family Income Percentile, Age 26	(4) Family Income Percentile, Age 26	(5) Family Income Percentile, Age 26
Normal	0.755^{+}	0.467	-0.0464	-0.00952	0.0773^{+}
	(0.434)	(0.808)	(0.460)	(0.0983)	(0.0436)
Observations	324	305	324	305	305
Weights	Unweighted	Precision Weights	Unweighted	Unweighted	Precision Weights
Scale	Per Childhood	Per Childhood	Per Childhood	Per Year	Per Year
Interpretation	Effect on Place	Effect on Place	Effect on Place	Effect on Person	Effect on Person

Standard errors clustered by state. $^{+}$ $p < 0.1$, * $p < .05$, ** $p < .01$. Outcome data in columns 1 and 2 are from Chetty et al. (2018), and outcome data in columns 3-5 are from Chetty and Hendren (2018).

Table A42 shows a similar analysis for marriage, and is an expanded version of Table 3 columns (5) and (6). As with the other two outcomes in Tables A40 and A41, the alternative weighting schemes which make the regressions in Columns (1) and (5) more comparable are also less powerful, and while the point-estimates are still positive, they are not statistically significant. In this case, the point estimate on permanent residents in Column (3) is also positive but not statistically significant.

Table A42: Effect on Marriage, 25th percentile parental income, Robustness

	(1) Married 2015	(2) Married 2015	(3) Married Age 26	(4) Married Age 26	(5) Married Age 26
Normal	1.662^{*}	1.011	0.501	0.138	0.0856^{+}
	(0.817)	(1.709)	(0.807)	(0.201)	(0.0470)
Observations	324	300	324	305	300
Weights	Unweighted	Precision Weights	Unweighted	Unweighted	Precision Weights
Scale	Per Childhood	Per Childhood	Per Childhood	Per Year	Per Year
Interpretation	Effect on Place	Effect on Place	Effect on Place	Effect on Person	Effect on Person

Standard errors clustered by state. $^{+}$ $p < 0.1$, * $p < .05$, ** $p < .01$. Outcome data in columns 1 and 2 are from Chetty et al. (2018), and outcome data in columns 3-5 are from Chetty and Hendren (2018).

D.5 Causal Effects on Children at the 75th Percentile of Parental Income

Table A43 is the same table as Table 3, but for children born to parents at the 75th percentile rather than the 25th percentile. For every outcome, the effect using the Chetty and Hendren (2018) data are insignificant.

Table A43: Causal Effects on Children: College Attendance, 75th percentile parental income

	(1) Some College Age 25+	(2) Attended College Age 18-23	(3) Family Income Percentile, 2014-15	(4) Family Income Percentile, Age 26	(5) Married 2015	(6) Married Age 26
Normal	0.775	0.0142	0.240	0.0206	0.648	0.0176
	(0.478)	(0.0486)	(0.201)	(0.0433)	(0.548)	(0.0636)
Observations	324	305	324	305	324	300
Weights	Unweighted	Precision Weights	Unweighted	Precision Weights	Unweighted	Precision Weights
Scale	Per Childhood	Per Year	Per Childhood	Per Year	Per Childhood	Per Year
Interpretation	Effect on Place	Effect on Person	Effect on Place	Effect on Person	Effect on Place	Effect on Person

Standard errors clustered by state. $^{+}$ $p < 0.1$, * $p < .05$, ** $p < .01$. Outcome data in column 1, 3, and 5 are from Chetty et al. (2018), and outcome data in columns 2, 4, and 6 are from Chetty and Hendren (2018).

D.6 Education Results by Race and Sex

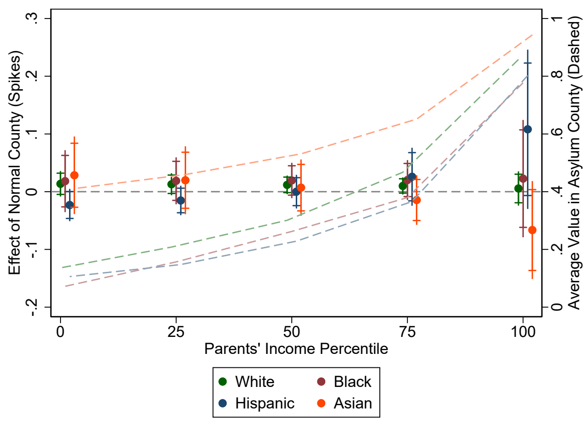
In this appendix, we present the effect of normal schools on educational attainment, by race and gender.

Note that the sample of counties is different across races due to data availability, making comparisons across race difficult. For example, there are 324 counties in the regressions comparing college attainment of White individuals in normal school versus asylum counties, but only 172 counties in the regressions for Black individuals.

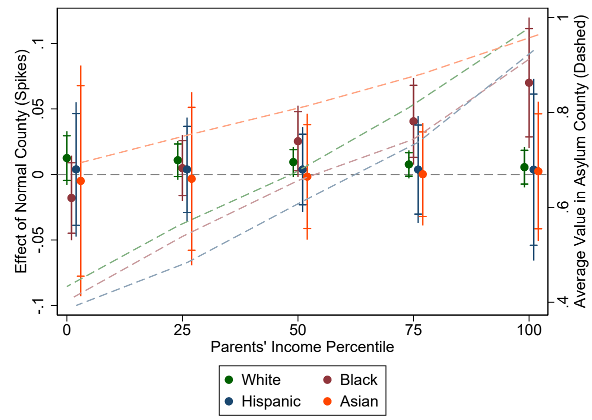
In Figure A10, we show the results from Figure 2, by race. The effects on high school attainment are very large for Hispanics, especially at the lower end of the income distribution. For at least some college, there is a large effect for Black children whose parents are at the top of the income distribution. And for college degrees, there is a large effect for Hispanic children with parents at the top of the income distribution. The results at the top of the distribution contrast with the results averaging across races being the least significant at the top of the distribution.

For Black individuals from higher-income families the effects on at least some college are larger in magnitude than the effects on four-year degree attainment (and they are statistically significant). If the increase in those with exactly some college (e.g., at least some college minus at least a four-year degree) were statistically significant, this could imply that regional public universities are inducing additional enrollment but completion rates are low. However, this increase is not statistically significant.

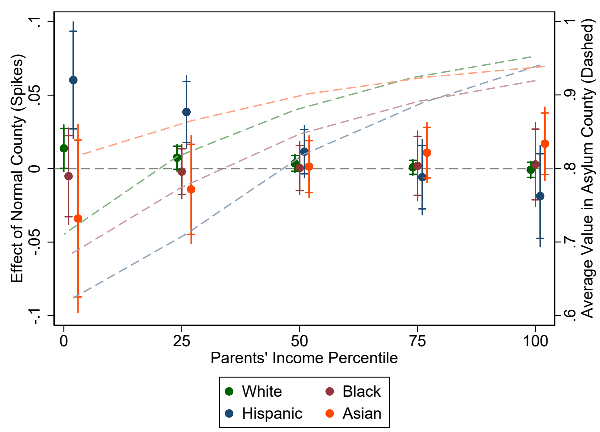
For sex, presented in Figure A11, the most interesting result is that across the income distribution, the effect on 4-year college degrees is stronger for women, though the confidence intervals overlap. For high school degrees, the result is slightly stronger for men, at least at the bottom of the income distribution, though again the confidence intervals overlap.



(a) At least 4-year College Degree, Age 25 and over

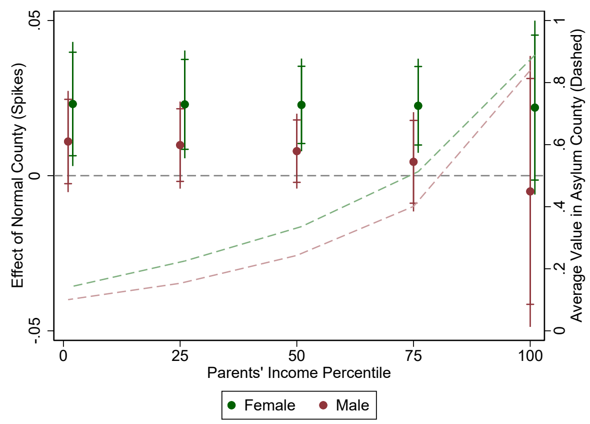


(b) At Least Some College, Age 25 and over

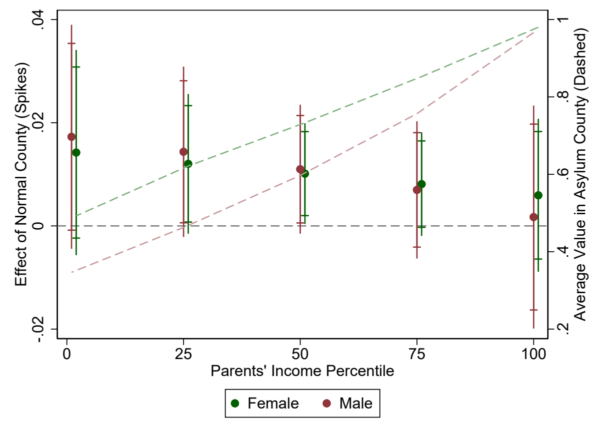


(c) At least HS Graduate or GED, Age 19 and over

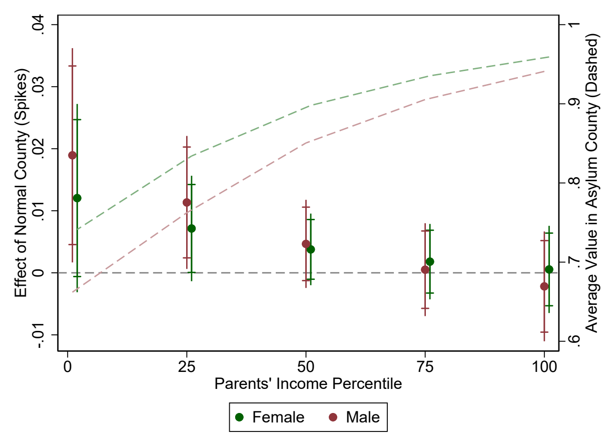
Figure A10: **Effect of a normal school on education, by race.** Dots with spikes represent the estimated effect of a normal school on the outcome variable, conditional on state fixed effects. Outcome variables are measured at different percentiles of the parents' income distribution, which is the x-axis. The spikes span the 95 percent confidence intervals, with the cross-bars at the 90 percent confidence intervals. Standard errors are clustered at the state level. The estimates and confidence intervals correspond to the y-axis on the left-hand side of the figure. The dashed lines show the mean of the outcome variable in asylum counties, and correspond to the y-axis on the right-hand side of the figure.



(a) At least 4-year College Degree, Age 25 and over



(b) At Least Some College, Age 25 and over



(c) At least HS Graduate or GED, Age 19 and over

Figure A11: **Effect of a normal school on education, by sex.** Dots with spikes represent the estimated effect of a normal school on the outcome variable, conditional on state fixed effects. Outcome variables are measured at different percentiles of the parents' income distribution, which is the x-axis. The spikes span the 95 percent confidence intervals, with the cross-bars at the 90 percent confidence intervals. Standard errors are clustered at the state level. The estimates and confidence intervals correspond to the y-axis on the left-hand side of the figure. The dashed lines show the mean of the outcome variable in asylum counties, and correspond to the y-axis on the right-hand side of the figure.

Table A44: Comparison of likely parents of migrant children in normal school vs. asylum counties, interacted with age

	(1) Highest Education	(2) Num. Adults Employed	(3) Log Parental Income
Normal	-0.235	-0.0327	-0.0775
	(0.142)	(0.0212)	(0.0623)
Age X Normal	0.00421	0.000910	0.000710
	(0.0115)	(0.00184)	(0.00361)
Observations	6089	6089	6079

Standard errors clustered by state

$^{+}$ $p < .1$, * $p < .05$, ** $p < .01$, *** $p < .001$

D.7 Is there evidence that children moving into normal school regions at later ages are different than children moving into asylum regions at later ages?

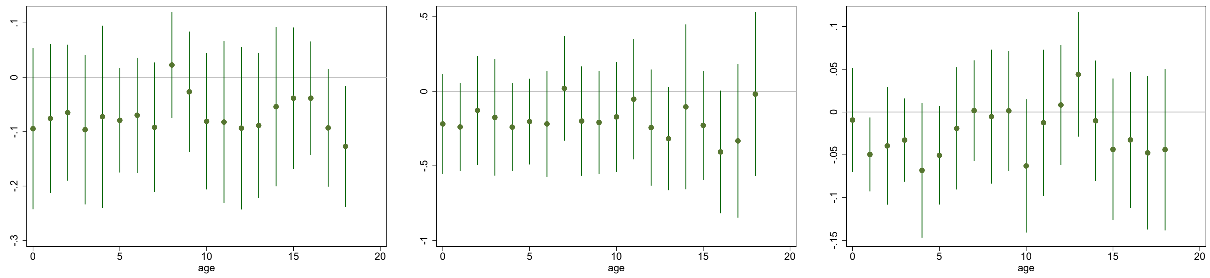
One threat to our analysis based on the causal estimates on individuals, from Chetty and Hendren (2018), is that it could be that the children who move into normal school counties at later ages are systematically different than those who move into asylum counties at later ages. In this appendix we test this hypothesis and do not find any evidence for it.

We use data from the 2005-2009 ACS to identify children of differing ages that moved into PUMAs with either a normal school county or asylum county in the last year.[63] We then look at characteristics of their parents to see if they differ on several important observable characteristics that may predict outcomes for the children. In particular, we investigate whether parents of older children have more education, more income, or more working adults, differentially in normal school relative to asylum counties.

It is important to note that we are not claiming that there are no differences in family characteristics based on the child's age when they migrate. Of course, migrant parents with older children are going to look different than migrant parents with younger children. We also are not concerned if there are overall differences in the migrants that move into normal school counties or asylum counties; this is the endogeneity that the Chetty and Hendren (2018) data is supposed to overcome. What would be a concern is if migrant parents with older children look different than migrant parents with younger children, and this difference depends on whether they move into a normal school or asylum county.

One issue with our approach is that the public ACS data do not identify the county reliably for our sample. So we we expand our geographic definition to be based on the public use microdata area (PUMA), which we can observe in the public data. We then assign households to normal school counties based on geographic crosswalks. For many counties, the PUMA exactly coincides with the county, and for many others, the PUMA is larger but

[63] 2005-2009 is a bit later than the children that are tracked in the Chetty and Hendren (2018) study. However, using the 1990 or 2000 census would require us to base these estimates on 5-year migration, which means we cannot precisely estimate the age at which the child moves.



(a) Log Household Income (Adults only)

(b) Years of education (max of adults in household)

(c) Number of Working Adults in Household

Figure A12: Comparison of likely parents of migrant children in normal school vs. asylum counties, by age of child. Each spike plots the coefficient and 95 percent confidence interval from the regression in equation (A2).

has only one normal school or asylum county in it. For a few, the PUMA encompasses both normal school and asylum counties. Unfortunately, this leads to bias in the direction of not finding significant differences in normal school areas, but we do not think there is a better approach with publicly-available data.

Limiting the sample to households that moved into the PUMA in the last year, we calculate county-level average parental characteristics separately for households with children at each age. For each county in our sample, we calculate the average of log income of the child's parents, the number of working parents of the child, and the average number of years of schooling of the child's most educated parent. For all these variables, we use the IPUMS classification of likely parents to identify characteristics of the parents.

We then estimate the regression

$$y_{ia} = \beta \text{Normal}_i \times a + \gamma \text{Normal}_i + \delta_{sa} + \epsilon_{ia}$$

where y_{ia} is the average value for households with children of age a in county i, a is the age of the child, and δ_{sa} is a state-age fixed effect. The unit of observation is at the county-age level, and we include children age 0 to 18.

The results are in Table A44. None of the coefficients on Normal $\times$ a are significant, which reassures us that we are not finding evidence of selection that would bias our results based on the Chetty and Hendren (2018) data.

We also estimate separate regressions by age, in case of a non-linearity that is not captured by our previous specification:

$$y_{ia} = \beta_a \times \text{Normal}_i + \delta_{sa} + \epsilon_{ia} \tag{A2}$$

A plot of the β_a are in Figure A12. There is no clear trend or visual evidence of differential effects by age.

E Details on The Freshman Survey Analysis

We merge student zip codes to counties using the CDC County Cross Reference File (Centers for Disease Control and Prevention, 1988). If the zip code is in a normal school and an asylum county, we assign the zip code to the normal school county. Roughly 3% of the observations that we classify as being from a normal school or asylum county reported a zip code that merged to both a normal school and an asylum county. If the zip code is in multiple normal, or in multiple asylum, counties we assign it to just one of the counties. If one of the counties matches the county of the respondent's university we choose that county. Roughly 3% of the students who grew up in normal school or asylum counties reported zip codes that merged to more than one normal or more than one asylum county.

We merge the public-access TFS data to the restricted-access TFS data in order to obtain the IPEDS ID of the university the student is attending (in the restricted-access data), by merging on TFS university code and year. A very small number of observations in the public-access data have TFS code-year pairs that are not in the restricted-access TFS data (roughly 3000 or .08%, coming from 62 TFS code-year pairs). Since we do not have the IPEDS ID for these students' universities, the variables denoting whether they attend a previous normal school, and whether they attend a university in their county will be missing. However, our main regressions focus on differences in reasons for choosing universities based on the student's home county. Thus, as long as the variable denoting their home county is not missing they will be included in these regressions.

There are roughly 33,000 observations from the public-access data for whom we do not obtain the county FIPS or other IPEDS variables for their university (roughly .9% of the sample of individuals who grew up in normal school or asylum counties). As discussed above, for roughly 3000 observations this is because the TFS university-year pair for the observation in the public data is not in the restricted data, so we cannot obtain an IPEDS ID for the university. For roughly 13,000, the TFS code-year pair is in the restricted-access data, but there is not an IPEDS ID associated with that pair in the restricted-access data. For the roughly 17,000 remaining observations, there is no IPEDS ID because the individual is attending a university that is not Title-IV eligible, or is in a state or geographic area without any normal schools or asylums in our data (Alaska, Hawaii, Puerto Rico, and the Virgin Islands).[64] For roughly 97% of these 17,000 observations, they are attending U.S. service academies (specifically the U.S. Military Academy, U.S. Coast Guard Academy, and U.S. Naval Academy). As discussed above, as long as these 33,000 observations have non-missing home county FIPS codes, they will be included in the main analysis which does not rely on the university the student attended. However, their universities will not be included in the regression to test differential participation of universities in normal school versus asylum counties, because these regressions are weighted by total bachelor's degrees awarded by the university that year. Because these universities are missing all IPEDS variables, they will be excluded from this regression. This TFS participation regression addresses the concern that we may be missing many individuals from some counties who stay close to home for college if their college does not participate in TFS. However, this is less of a concern for the U.S. service academies given that applicants to these academies must be nominated to apply, and

[64] We use a roster of Title IV eligible institutions in IPEDS.

there are geographic restrictions on the number of students from each congressional district in the U.S.[65]

For the roughly 30,000 individuals whose TFS code-year pair is in the restricted access data, but missing data from IPEDS, we obtain their institution's county using the institution's zip code in the restricted data. We merge zip codes to counties using a similar procedure as that described for student zip codes above. For these observations, if the institution zip code merges to a normal school and asylum county we keep the observation from the normal school county. If any of the counties that merge to an institution's zip code are the same as any of the counties that merge to the student's zip code, we use that institution and student county. By using the institution zip code from TFS restricted data, rather than IPEDS, we are able to include some additional observations when looking at the outcome of whether a student attended a university in their home county. There are roughly 600 individuals who are in the restricted and public TFS data, but missing data from IPEDs and missing the institution zip code from TFS. These individuals attend five different universities, and we obtain the university's county using the institution's city and state in TFS. We look up the county name using the City-to-County Finder (StatsAmerica, 2023), and then obtain the county FIPS code using U.S. Census Bureau (2002).[66] Further, for the roughly 17,000 observations attending institutions that are not in our IPEDS roster, we are able to impute that they are not attending a previous normal school, because all of the universities that started as previous normal schools are in the IPEDS roster.

The TFS includes survey weights. These weights are at the institution-type by gender by year level, and give more weight to students from types of universities that were less likely to participate in TFS in a given year. The weights bring the "male and female counts up to the national number of first-time full-time freshmen in each stratification cell" (Pryor et al., 2010, p. 142). The weights are based on stratification groups based on institutional race (e.g., predominantly non-black vs. predominantly black), type (e.g., four-year college, university), control (e.g., public, private nonsectarian, Roman Catholic and other religious) and selectivity. The exact groups differ slightly over the sample period. The TFS weights in our dataset also assign zero weight to students at nearly all two-year colleges, at institutions with low student response rates (in 2010 they report the threshold was 65% for four-year universities and 75% for four-year colleges), and to part-time students, and non-first-time college students (Pryor et al., 2010). There are a total of roughly 2.5 million individuals in our sample who grew up in normal school or asylum counties and have positive TFS weights.

When merging TFS data with IPEDS enrollment data, a small fraction of universities have response rates greater than one. This may be due to differences between the enrollment measure in IPEDS (first-time undergraduate degree/certificate-seeking students) and the set of students to whom the university administers the survey.

[65]Individuals attending the U.S. Air Force Academy did get merged to the university's county FIPS code even though it is not Title-IV eligible. However, this institution is not in a normal school or asylum county, and so also will not be included in the regression testing differential participation of universities in normal school versus asylum counties.

[66]For one city, O'Fallon, MO, the tool StatsAmerica (2023) does not yield the county name, and so we obtain it from National Association of Counties (2023).

E.1 Differences in TFS Participation Between Universities in Normal School and Asylum Counties

In this section we provide more details on our test for whether universities in normal school counties are more likely to respond to TFS than universities in asylum counties. We note that our main specification addresses many concerns about differential participation in TFS by utilizing the TFS weights, which were designed to address differential TFS participation by university type.

Using IPEDS, we construct a dataset of all four-year, Title-IV-eligible universities in normal school and asylum counties in each year (using the university's county FIPS code). We construct a separate dataset of all the universities that respond to TFS in each year, with students that have positive TFS weights. To do this, we first merge the full public data (not limited to students who grew up in normal school or asylum counties) to the restricted data with IPEDS ID, merging on TFS ID and year, keeping only observations with positive TFS weights. We then collapse at the university-year level to obtain a dataset with the universities responding to TFS in each year. We then merge this to the roster of all universities in normal school and asylum counties in each year, and keep the universities located in normal school and asylum counties.[67] The university-year observations in the full IPEDS roster that merge to TFS data are those that respond to TFS that year. The remaining university-year observations do not respond to TFS.

We test for within-state-year differences in TFS participation between universities in normal school counties and universities in asylum counties.[68] Specifically, we estimate:

$$\text{TFS participation}_{jt} = \beta \text{Normal}_j + \alpha_{st} + \epsilon_{jt}$$

We weight observations by the total number of bachelor's degrees awarded by the university in each year. This weighting incorporates that we should be less concerned if small universities in asylum counties are not participating in TFS, as this is less likely to bias our results. Similarly, we may also be less concerned if online universities in asylum counties are less likely to participate in TFS, as these universities may be less likely to enroll local students. We estimate an additional specification in which we exclude universities in any year for which at least 50% of the university's enrollment in 2018 was enrolled in distance education. The distance education variable is not available in earlier years. Roughly 13% of the universities in this specification did not merge to the 2018 roster, and for the purposes of this specification we assume they were not "distance enrollment" universities.

We cluster standard errors at the county level. We find there is no statistically significant

[67] As we discuss above, this will drop universities in the public TFS that were not merged to an IPEDS ID. However, as noted above, these numbers were small.

[68] A reader may also be interested in the raw averages by type of county. For universities in normal school counties, the mean likelihood of TFS participation and students with positive TFS weights, over the years from 1982-2010, is 16%, while in asylum counties it is 13%. Weighted by total bachelor's degrees awarded in each year, these means are 23% in both normal school and asylum counties. Not limiting to universities with positive TFS weights, the weighted means are 36% in each type of county. There are 379 universities across 168 normal school counties, and 198 universities in 66 asylum counties, that respond to TFS from 1982-2010. There are 308 universities with positive TFS weights across 136 normal school counties, and 155 universities in 62 asylum counties, that respond to TFS from 1982-2010.

difference in TFS participation in a given year between universities in normal school counties and same-state universities in asylum counties (Table A45).[69]

Table A45: University Participation in TFS: Universities in Normal School vs. Asylum Counties

Y = Respond to TFS	(1)	(2)
Univ. in Normal School County	-0.042	-0.042
	(0.045)	(0.045)
Sample	All	Exclude distance education
Observations	22,959	22,306
R-squared	0.140	0.142

Notes: $^{+}$ $p < 0.1$, * $p < .05$, ** $p < .01$. Observations are at the university-year level. All regressions include state-year fixed effects. Column 2 excludes universities in any year for which at least 50% of the university's enrollment in 2018 was enrolled in distance education. Observations are weighted by the total bachelor's degrees awarded by the university in that year. Standard errors are clustered at the county level. See text for details.

E.2 Student's Reported Zip Code

We identify students from normal school and asylum counties using the zip code they report on the survey. Through 2000, students were asked for their address at the top of the survey, including their zip code. Starting in 2001, students were asked for their "permanent/home address," including their zip code. In this section we address the possibility that students report their address at the university rather than their family's address, which we are using as a measure for where the student grew up.

If students are reporting zip codes for their residence in college, we would expect many of the reported zip codes would match the university's zip code. This does not seem to be the case. Of the people with positive TFS weights whom we classify as growing up in a normal school or asylum county, based on their reported zip code, only 3% are reporting a zip code that is the same as their university. This suggests students are not filling out their address using their residence in college.

In addition to asking students for their address, the survey separately asks students for the distance between the college and their "permanent home." As a second test, we compare the zip codes reported to the separate question on home-university distance. If the people who report the same zip code as their university (or a zip code in the same county as their university) are actually reporting their college residence, we would not expect them to report home-university distances that are closer relative to the full sample of students we

[69] The point estimate is -.042 with a standard error of .045. In the regression sample, the weighted mean of the dependent variable in asylum counties is .234. We focus on showing no differential participation between universities in normal school and asylum counties. When we look at students' differential likelihood of attending college close to home (equation 1), we include state fixed effects. Thus, participation of farther universities, for example in other states, should matter more similarly for students growing up in normal school and asylum counties.

have classified as being from normal school or asylum counties. We do this comparison for people whose reported home zip code is in the same county as their university, as well as for people whose reported zip code is the same as their university.

Among all the students we have classified as being from normal school or asylum counties, and who have positive TFS weights, roughly 11% report the university is less than or equal to 10 miles from their permanent home and 23% within 11-50 miles. Among students whose reported zip code is in the same county as their university, and who have a non-missing response to the home-university distance question (roughly 23% of the sample), those percentages are 40 and 48%.[70] This suggests that for most students, they are reporting the zip code where they grew up rather than where the university is located. Some may report their residence in college, but this fraction appears small. Only roughly 73,000 students report a zip code that is the same as their university's zip code, and have a non-missing response to the home-university distance question, and positive TFS weight. Of those, 39% report their permanent home is within 10 miles of the university, and an additional 8.4% report it is within 11-50 miles. While roughly 50% report farther homes despite reporting a zip code that is the same as their university's, and this is potentially consistent with reporting college residence, the overall number of these individuals is low (roughly 1.6% of the overall sample of people we have classified as having grown up in normal school or asylum counties, and who have non-missing responses to the home-university distance question and positive TFS weights).

Finally, we test whether there is a discontinuous change in the results in 2001 consistent with the survey asking for "permanent/home address" instead of simply for the "address." We run the following regression:

$$\text{Attending university within 10 miles of home}_{it} = \beta_t \text{Normal}_{it} + \alpha_{st} + \epsilon_{it}$$

where an observation is an individual answering in year t, but we weight it by one divided by the total number of observations for that county. α_{st} is a state-year fixed effect. The weighting means that if we replace β_t with β and α_{st} with α_s, i.e. removing the time dimension from the regression, then the results would be identical to our main specification. We cluster standard errors at the level of the student's home state. The results with the time dimension are in Figure A13, and there is no obvious jump in 2001.

E.3 More Questions from the Freshman Survey

In this section, we show the differential answers from many more questions in The Freshman Survey. As in Section 3, we regress the answer to the question on a dummy for having a home zip code in a normal school county, with state fixed effects (regression specification (1)).

In Table A46, we show the likelihood of going to a college in various distance bins. We

[70] As we note above, a very small fraction of zip codes are in multiple counties. For this exercise, for each individual we merge to all the counties associated with their home zip code. We then determine whether any of these counties matches the university's county, and obtain the distribution of home-university distance for those individuals whose reported zip code is in their university's county (using just one observation per individual).

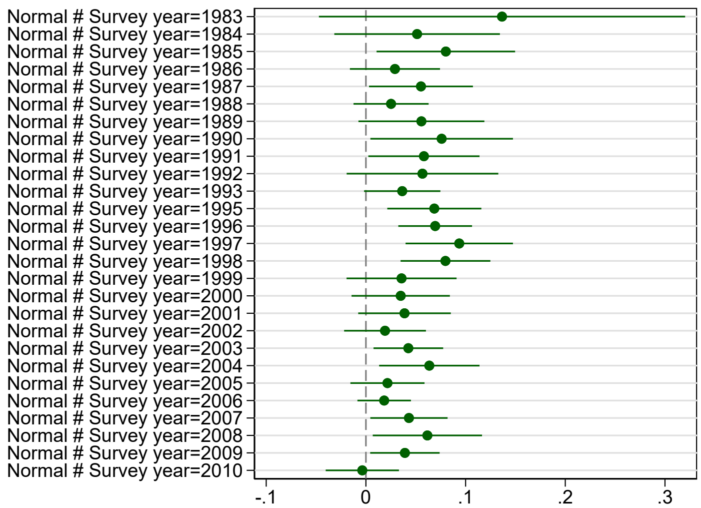


Figure A13: Differential probability of going to university within 10 miles, students from normal and asylum counties, conditional on state, by year

previously showed the first column in Table 4, but we show all the other bins of distance asked in the survey here for completeness. We see that students from normal school counties are less likely to be attending universities between 10 and 100 miles away. Other distances are not statistically significant.

Table A46: Differential Likelihood of Attending a College that is Various Distances from Home, for Children Growing Up in Normal or Asylum Counties

	Home-University Distance (miles)				
	(1)	(2)	(3)	(4)	(5)
	$\leq$10	11-50	51-100	101-500	>500
Grew up in normal school county	0.0462**	-0.0303*	-0.0302*	0.0163	-0.00194
	(0.0131)	(0.0148)	(0.0116)	(0.0139)	(0.00714)
Observations	324	324	324	324	324
R^2	0.283	0.231	0.255	0.225	0.630

Standard errors clustered by state. * p<0.05, ** p<0.01, *** p<0.001

Observations are at the county level. All regressions include state fixed effects.

In Figure A14, we show the results from many more questions that were asked in The Freshman Survey. As in Figure 5, the point is the regression coefficient from (1), and the spikes show 90 and 95 percent confidence intervals. We manually chose the questions from the survey that we thought would reflect differences in the composition of freshman from normal school and asylum counties, as well as questions about their academic preparedness and application behavior. We split the questions into eight categories: cost factors, location factors, job factors, college applications, information, academic factors, demographic characteristics, and other. In general, many of these questions are similar to ones in the main text, so these are presented to show that the results are robust to these other measures. In particular, for questions that ask students to rank the importance of various factors, we now include an indicator variable for just "very" important as an outcome variable. In the main text, we looked only at an indicator variable for "very" or "somewhat" important.

E.4 Robustness to Year Fixed Effects

In this section, we explore whether aggregating survey answers across years meaningfully changes our results. To do this, we take advantage of the individual-level data, and run the following regression where i is an individual instead of a county:

$$y_{it} = \beta \text{Normal}_i + \alpha_{st} + \epsilon_{it}$$

Using individual-level data allows us to compare students to other students from the same state and same year but in a different county. To make these results otherwise comparable, we weight the individual-level observations by the inverse of the number of observations we observe in that county over all years, multiplied by the TFS weights.[71] Under this weighting scheme, if we only include state fixed effects, the results are identical to our main specification. As before, we cluster standard errors by state.

Our results are in Table A47 and Figure A15. The results are not meaningfully different than the results we presented in Table 4 and Figure 5.

Table A47: Differential Likelihood of Attending University Close to Home

	(1) Attend Univ. within 10mi	(2) Attend Former-Normal within 10mi	(3) Attend Univ in county	(4) Attend Former-Normal in county
Grew up in normal school county	0.0525***	0.0717***	0.131***	0.165***
	(0.0122)	(0.0106)	(0.0278)	(0.0221)
Observations	2483010	2477038	2531208	2525784

Standard errors clustered by state. * p<0.05, ** p<0.01, *** p<0.001

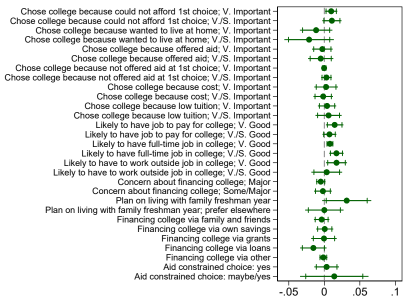
Observations are at the individual level and weighted so each county has equal total weight. All regressions include state-year fixed effects.

E.5 Splitting by Parental Income

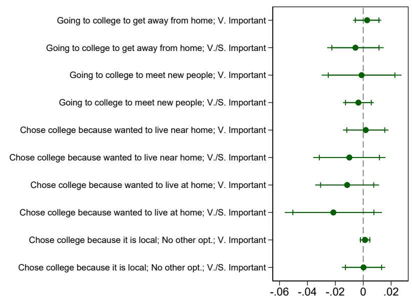
A reader might be interested in whether our main results from the TFS analysis differ by parental income.

To perform this analysis, we split our sample by the median income of self-reported parental income in each year. In particular, students were asked to report a bin of their parents' income. We found the median bin, and classified everyone strictly below the median bin as lower-income. Then we estimated the regressions separately for the high- and low-parental income students. The results are generally fairly similar for both groups, evidenced in Figure A16.

[71] An alternative approach would have been to collapse the data to the county-year level and run this regression without weighting. However, there are many county-years with very few observations and so the results are noisier.

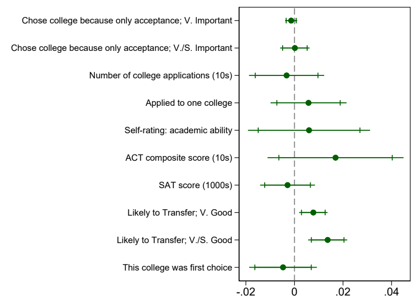


(a) Cost Factors

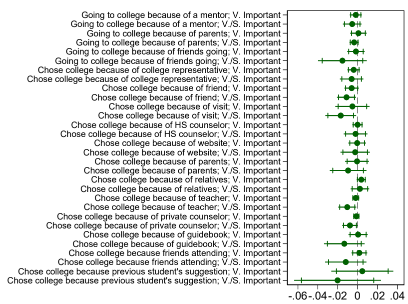


(b) Location Factors

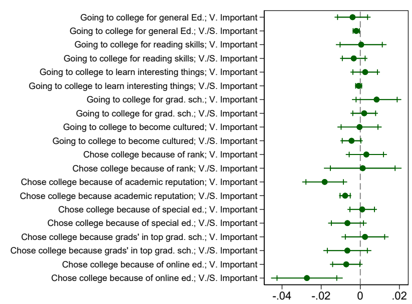
(c) Job Factors



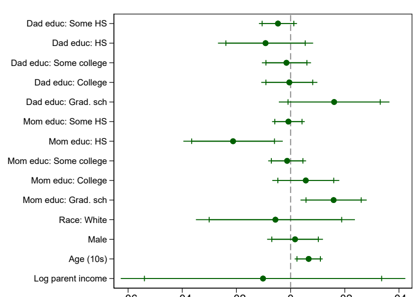
(d) College Applications



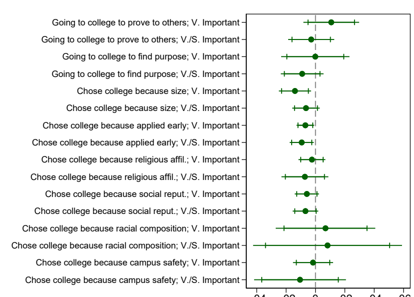
(e) Information



(f) Academic Factors

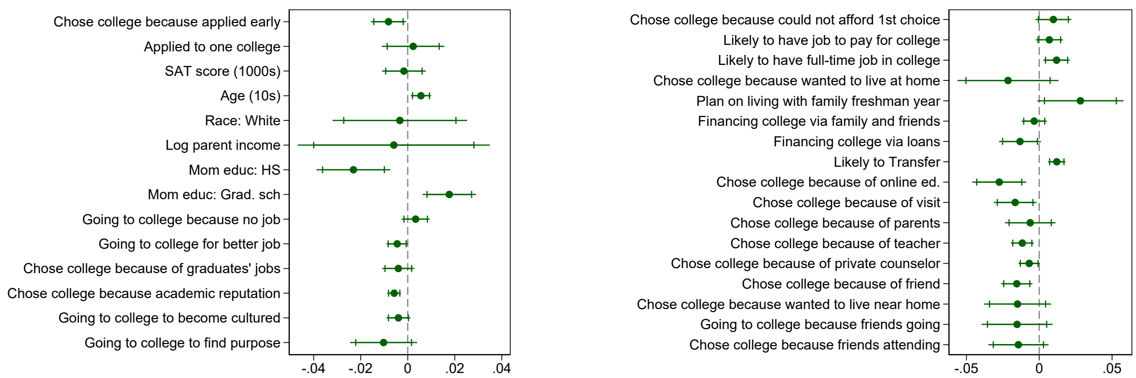


(g) Characteristics



(h) Other

Figure A14: The difference in Freshman Survey answers between college freshman who grew up in normal school counties versus asylum counties. For some questions that were on a three point scale, we report the difference in students reporting the top choice ("Very") or the top two choices ("Very" and "Somewhat"). These are abbreviated as "V." and "V./S." in the above figure. Spikes represent 95 percent confidence intervals and cross-hatches are at the 90 percent confidence intervals. Standard errors clustered by state.

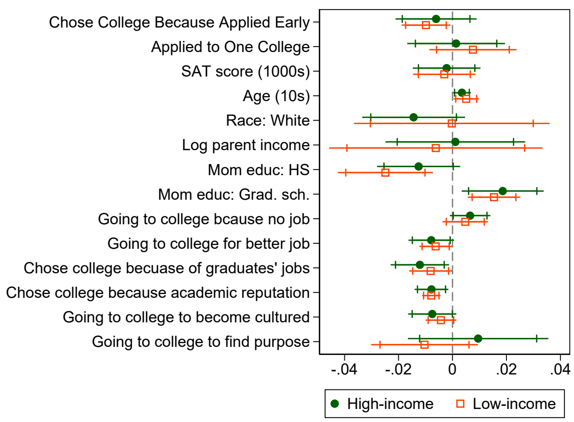


(a) Characteristics of Students

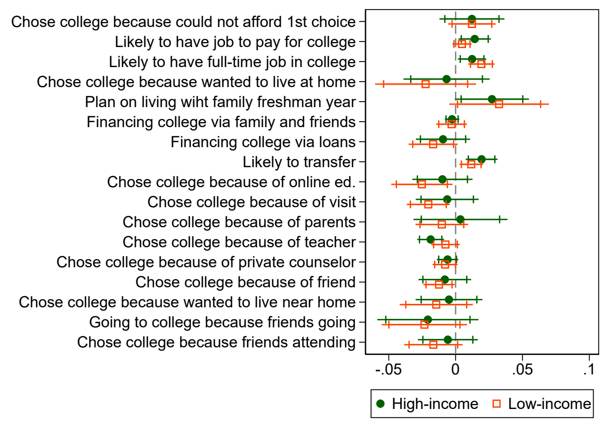
(b) Geographic Frictions of College Attendance

Figure A15: Differential answers to The Freshman Survey by students who grew up in normal school counties relative to same-state same-year students who grew up in asylum counties. For questions that are answered on a five point scale, we create a dummy variable if the student answered that the reason was “Very important/good” or “Somewhat important/good.” Spikes are 95 percent confidence intervals and cross-hatches are 90 percent confidence intervals. Standard errors clustered by state.

As we had discussed in the paper, the estimates from the TFS analysis potentially reflect effects of proximity to a historical normal school on who chooses to attend college, as well as effects of proximity on the characteristics of the children. Both of these may differ by parental income levels, making interpretation of the differences by parental income more difficult. Further, Figure 2 showed effects on college attainment even above the 50th percentile of the parental income distribution.



(a) Characteristics of Students



(b) Geographic Frictions of College Attendance

Figure A16: Differential answers to The Freshman Survey by students who grew up in normal school counties relative to same-state students who grew up in asylum counties. For questions that are answered on a five point scale, we create a dummy variable if the student answered that the reason was "Very important/good" or "Somewhat important/good." Spikes are 95 percent confidence intervals and cross-hatches are 90 percent confidence intervals. Standard errors clustered by state.