Abstract: Does the college major premium reflect returns to prior abilities or college education? We decompose the college major premium into labor market returns to multidimensional abilities (grit, interpersonal, and cognitive) and skills learned in college. This allows us to quantify how much of the college major premium is due to sorting on multi-dimensional abilities and how much is due to the differential labor market value of major-specific skills. We find that sorting on abilities accounts for 10-50% of the college major premium. We also provide novel estimates of complementarities and interaction effects between abilities and skills, since the returns to abilities vary significantly across college majors. We document that 40% of students who enter STEM degrees change major or drop out. We evaluate counterfactual policies to promote STEM degrees, accounting for the fact that many who start STEM degrees do not finish.

JEL: C32, C38, D84, I21, J24, J31.

Keywords: College Major Choice, Cognitive and Non-cognitive Abilities, Educational Investment.

*University of Chicago. Email: johneric@uchicago.edu
†University of Chicago, Stockholm School of Economics, and IZA. Email: jjoensen@uchicago.edu
§Arizona State University. Email: gregory.veramendi@asu.edu. We gratefully acknowledge financial support from the Swedish Foundation for Humanities and Social Sciences (Riksbankens Jubileumsfond) grant P12-0968 enabling the data collection for this project. The usual disclaimers apply.
1 Introduction

College major choices have widespread economic consequences. Income differences between fields of study ("college major premium") are as large as the differences between different education levels ("college premium") (Christiansen et al., 2007; Altonji et al., 2012; Kirkebøen et al., 2016) and have been increasing over time (Gemici and Wiswall, 2014; Altonji et al., 2014). Even if the college major premium is well-documented (Berger, 1988; Grogger and Eide, 1995; Paglin and Rufolo, 1990; Arcidiacono, 2004; Altonji et al., 2015) it is still not well understood what it embodies. Does it reflect sorting on abilities, differential labor market returns to abilities, or is it a causal effect of the skills acquired in the college major? In this paper, we decompose the college major premium into self-selection on multidimensional abilities (grit, interpersonal, and cognitive), sorting based on differential labor market gains to these abilities, and the causal effect of college major-specific skills. This allows us to quantify how much of the college major premium is due to self-selection, sorting on gains, and a causal effect of the college major. We also do counterfactual policy experiments to quantify how college major choice and graduation rates are affected by: (i) investments in pre-college abilities and (ii) providing incentives to enroll in particular majors.

We estimate a multistage sequential choice model of educational choices where we approximate decisions at each stage. In the first stage, high school graduates choose whether to enroll college or not. If they enroll in college, they choose which level and field to enroll in: two levels and nine fields. In the second stage, they choose whether to switch level and/or field, or to continue in the initially chosen level and field. In the third stage, they either dropout or graduate with a degree in the field and level chosen in the second stage. In the final stage, they work in the labor market and their earnings are determined by their initial abilities and their college (level and major) choices. We account for measurement error and biases in the measurement of three dimensions of latent ability: grit, interpersonal, and cognitive. This model allows us to isolate the effects of latent abilities such that we can decompose the college major premium into components consisting of: skills learned in college and sorting on and returns to multidimensional
abilities. The model also allows us to characterize the population at each decision node and who will be affected by a potential policy. This enables us to answer questions such as: What are causal returns at each decision node and how do they depend on abilities? Do individuals choose majors where they have highest gains to abilities? What is the effect of policies that encourage Science, Technology, Engineering, and Mathematics (STEM) enrollment? Are the skills that students are screened on upon college enrollment also those that translate into successful completion of majors – particularly STEM degrees – and are the same skills rewarded in the labor market for these fields?

We use a comprehensive administrative dataset of the whole Swedish population for our analysis. We observe detailed measures of abilities at age 16 in the 9th grade registry – the end of compulsory schooling – and age 18 from the military enlistment archives. We observe detailed educational codes for all enrollment spells, accumulated course credits, and acquired degrees in the Higher Education registry. We focus on the sample of high school graduates in 1993-94 and merge the measures of abilities and educational choices to their yearly earnings which we observe until 2013 – when they are in their late 30s. We are also able to link children to their parents and observe a rich set of background variables from the medical birth records and other administrative registers.

Our preliminary results suggest that endowments strongly influence both educational choices and outcomes: (i) Students notably sort across majors. Grit, interpersonal and cognitive abilities vary considerably across majors, and accounting for this selection is paramount. (ii) Controlling for initial abilities can explain 10 to 50 percent of the earnings differences between majors. (iii) The labor market returns to ability are quite heterogeneous after conditioning on field and level of education. For example, the labor market rewards Business majors for all three dimensions of ability, while Law majors are rewarded for higher interpersonal and grit ability, but not cognitive ability. In contrast, earnings of those with Education, Social Sciences and Engineering majors depend on interpersonal ability, but not on the other two dimensions of ability. (iv) We also show heterogeneity in the causal returns to major by abilities and can distinguish returns to enrollment and to graduation. Allowing for the option to change major is key to understanding the returns
to major at enrollment.

Our approach adopts a middle ground between the reduced form and structural literature.\(^1\) The three main advantages of this approach are: First, we estimate causal effects at clearly identified margins of choice. Second, we can estimate causal effects for populations affected by different policies. This allows us to pin down the characteristics of individuals affected by various policies. Third, we can decompose ex-post returns into both the direct causal effect and the continuation value.

There is a growing quasi-experimental and structural literature on college major choice.\(^2\) We bridge and extend the literature by taking the sequential nature of educational choices into account while allowing a year of education and individual abilities to have multiple dimensions: cognitive and two non-cognitive (grit and interpersonal). First, we show that accounting for the sequential nature of educational choices by allowing individuals to switch majors is important as the returns to enrolling in a college major are very different than the returns to graduating. The early literature considers returns to graduation (Berger, 1988; Altonji, 1993), while the recent quasi-experimental literature evaluates the causal returns at the time of college application (Kirkebøen et al., 2016; Hastings et al., 2013). The main advantage of the quasi-experimental approach of Kirkebøen et al. (2016) is that correction for selection bias is achieved through random variation in who gets accepted to college around clearly defined admission score cut-offs. They show how this type of random variation coupled with data on a rank-ordered list of majors and universities can identify the return to each major relative to the next-best major. They also show that estimated returns to major vary substantially by next-best major. We do not have access to application data, but our model delivers ranking of majors – both at the time of enrollment and at graduation. We find the distinction between causal returns at enrollment and graduation to be an economically important one. Second, our model not only delivers a preference ranking of majors, but provides a

\(^1\)We extend Heckman et al. (2016) to not only include binary sequential choices of levels of education, but also multinomial choices of fields of college major. Furthermore, our rich measures of skills allow us to include not only one latent factor of non-cognitive skills, but two factors – grit and interpersonal – which both matter at different margins.

\(^2\)This literature has recently been thoroughly review in Altonji et al. (2012) and Altonji et al. (2015).
cardinal measure of how close individuals are to indifference between majors. This cardinality obviously relies on the structure we impose on the choice problem and preferences, but is important for policy evaluation and to pin down who is affected by policies. This allows us to characterize the compliers to each type of policy; e.g. who is affected if the government provides very strong incentives to enroll in a STEM major. We show that who and how many are close to being indifferent to their first-best major (in absence of the policy) is an important driver of how strong the policy response is at the STEM graduation margin. Furthermore, we are able to estimate how the returns to the compliers compare to the average returns. Important for the identification of these channels is the fact that those who dropout, switch, and stay are characterized by different sets of abilities. Students often enroll in college with optimistic expectations about graduating with a STEM major (Arcidiacono et al., 2015), but almost half of those who initially intend to graduate with a STEM major either switch to another field or drop out (Altonji et al., 2015). High switching costs and dropout rates may be both privately and socially costly, and our model can quantify their economic importance. Third, common to both the quasi-experimental and structural papers is that abilities (or comparative advantage) is one-dimensional.\(^3\) We show that also taking non-cognitive abilities into account is paramount in order to better explain why individuals choose different majors. For example, why do so few choose STEM majors despite the high premium? One reason could be that these majors require both high cognitive ability and high grit. The structural literature – most notably Arcidiacono (2004) – has many of the advantages listed above in that it does take the sequential nature of educational decisions into account. The literature finds that preferences are very important determinants of major choices, but sorting is only allowed on one-dimensional cognitive abilities.\(^4\) This conclusion may

---

\(^3\) Multiplicity of abilities has been shown to be important for a range of labor market, education, social, and health outcomes in a large body of work; see e.g. Heckman et al. (2006); Lindqvist and Vestman (2011); Heckman et al. (2014, 2016).

\(^4\) Arcidiacono (2004) concludes that virtually all cognitive ability sorting is because of preferences for particular majors in college, but also shows that those who perform worst than expected are more likely to switch to lower paying majors while those who perform better than expected are more likely to switch to higher paying majors. Learning about academic ability has been found to be important for college major switching and dropout decisions (Zafar, 2011; Stinebrickner and Stinebrickner, 2013) and students have also been found to respond strongly to new information on major-specific returns (Wiswall and Zafar, 2015). We focus on self-selection and sorting based on multidimensional abilities, but also allow
reflect sorting on initial and persistent non-cognitive abilities, but is interpreted as preferences by the existing major choice literature. This relates to the conceptual issue whether non-cognitive abilities (e.g. grit and interpersonal) most accurately describe abilities or preferences.\textsuperscript{5} We are the first to analyze sorting on multi-dimensional abilities and show that sorting patterns are more complex than previous papers suggest.

College major choices are socially and economically important because they determine the skill composition of the workforce. Thus may be important drivers of a countries’ development, innovation, economic growth, and competitive advantage. Our model allows us to perform counterfactual policy simulations to assess the impacts of: (i) early childhood programs that increase initial abilities and (ii) providing grants to individuals conditional on STEM enrollment. We quantify how such policies have very different impacts on (STEM) enrollment and (STEM) graduation rates.

Knowing how much of the income dispersion across fields of college major is driven by initial abilities relative to the skills acquired in college is pivotal in order to design better education policy and to know when to target resources. For example, many argue there is a scarcity of STEM skills.\textsuperscript{6} If so, it may improve efficiency to target resources towards investments in STEM skills. If the STEM premium mainly reflects initial abilities of those who choose the STEM fields, then resources should be targeted earlier in the life-cycle. Analyzing self-selection and sorting on gains to multiple abilities is important in order to know which types of early interventions are most beneficial. If the premium mainly reflects differences in grit and interpersonal skills, then potential policy interventions depend on the extent to which these skills are malleable at later ages.

\textsuperscript{5}Bowles et al. (2001) argue that character traits like persistence or dependability could be viewed as incentive-enhancing preferences which employers value in the face of incomplete labor contracts.

\textsuperscript{6}See e.g. Altonji et al. (2012), Altonji et al. (2015), adn the references therein.
2 Institutional Setting

After nine years of compulsory schooling, most Swedish students enroll in high school. We focus on those who graduate from high school in 1993-94. These comprise the pool of potential college applicants. Meeting the basic requirements for college enrollment requires completing three years of academic high school or two years of vocational high school followed by a year of an intensive college preparatory courses. College admission is predominantly conditional on high school grade point average (GPA), but other factors also affect the admission score; including high school track and course choices, labor market experience, and the Swedish Scholastic Aptitude test (SweSAT). College admission is centrally administered and a college application consists of a rank-ordered list of college-major choices. Higher education is tuition-free for all students and largely financed by the central government. College students are eligible for universal financial aid of which around one third of the total amount is a grant (or scholarship) and the remaining two thirds are provided as a loan. Student aid is largely independent of parental resources, but means-tested on student income and the maximum eligibility period is 240 weeks; i.e. the equivalent of 12 semesters or six enrollment years. Student loans are subsidized and the loan repayment plan was income-contingent for those in our sample.

3 Data

We merge several administrative registers via a unique individual identifier for the population of Swedes born in 1965-83.

Our measurements of health, abilities, and family background come from the Medical Birth Registry that is administered by the National Board of Health and Welfare (Socialstyrelsen), the Military Enlistment archives administered by the Swedish Defence...
Recruitment Agency (Rekryteringsmyndigheten) as well as several registers administered by Statistics Sweden (SCB).

The Medical Birth Registry contains measures of the child’s in utero environment and health status at birth; incl. maternal diagnosis and complications during pregnancy and delivery, child birth weight, indicators for whether the child is heavy or light for gestational age, APGAR score (Apgar, 1952) at 1, 5, and 10 minutes after birth, and child child diagnosis at birth for the cohorts born in 1973-83.

The Military Enlistment archives contain cognitive test scores, psychological assessments, health and physical fitness measures collected during the entrance assessment at the Armed Forces’ Enrollment Board. The enlistment was mandatory for all Swedish males at age 18 until 2010, thus for all males in our sample who are Swedish citizens. The entrance assessment spans two days. Each conscript is interviewed by a certified psychologist with the aim to assess the conscript ability to fulfill the psychological requirements of serving in the Swedish defense, ultimately in armed combat. The set of personal characteristics that give a high score include persistence, social skills, and emotional stability.

To sharpen our interpretation of the latent ability factors, we merge these registers to the ETF surveys administered to 3rd, 6th, and 10th grade students by the Department of Education and Special Education, Gothenburg University. This survey was administered to random samples of four cohorts in our population (1967, 72, 77, and 82) and includes extensive measures of aptitude and achievement tests, absenteeism, special education and tuition, and grades in various courses through compulsory schooling, as well as extensive student and parent surveys related to student achievement, grit, confidence, inputs, and interpersonal skills.

We also have detailed data on educational choices and outcomes from the 9th grade registry (incl. grades in individual courses and GPA), the High School registry (incl. grades in individual courses, GPA, track and specialization choices), and the Higher Education registry (incl. detailed educational codes for all enrollment spells, course

\[ \text{Härnqvist (1998) provides additional details on the construction of the survey.} \]
credits accumulated during enrollment, and acquired degrees). College applicants are screened based on their high school course choices and GPA. Some of them are also admitted based on high performance in the Swedish Scholastic Aptitude Test (SweSAT) on which we have overall test scores and sub-scores on every attempt through the Department of Applied Educational Science at Umeå Universitet.

From the Higher Education registry, we observe the level and field of every enrollment spell. We classify all spells into two levels ($\leq 3$ years; $\geq 4$ years) and nine fields (1. Education; 2. Humanities and Art; 3. Social Sciences and Services; 4. Math, Natural, Life and Computer Sciences; 5. Engineering and Technical Sciences; 6. Medicine; 7. Health Sciences, Health and Social Care; 8. Business; 9. Law).

The Multigeneration registry allows us to link children to their parents. It also contains information on family size and composition. Additional background variables are obtained from the longitudinal integration database for health insurance and labour market studies (LISA) from which we have yearly observations during the period 1990-2013. This allows us to observe individual employment status and earnings when they are 30-48 years old and parental background variables (incl. age, civil status, highest completed education, employment, earnings, and disposable family income). We supplement this with earnings information from the Registerbased Labor Market Statistics (RAMS) for the years 1986-89 and information on disposable family income from the Income and Tax registry (IoT) for the years 1978-89. This means that we can measure disposable family income of parents (parental earnings) from birth (age 3) to age 30 years old for the youngest cohort and from age 13 (age 16) to 48 for the oldest cohort in our sample.

3.1 Sample Selection

We focus on males, since military enlistment at age 18 was only mandatory for Swedish males and these scores are important measures for our factor model. We select a sample of high school graduates in 1993-94. The reason we focus on these cohorts is twofold: First, the detailed college credit data only exists form 1993 onwards and this is also the year the classification of higher education in Sweden changed considerably. Second, the four
sub-scores for the cognitive test taken at military enlistment are only sparsely observed for those who are born after 1976.

4 Empirical Model of Major Choice and Earnings

This paper estimates a sequential model of schooling decisions and labor market outcomes. The decision tree of this model is illustrated in Figure 1. Our analysis focuses on the population of individuals who have at least a high school degree. High school graduates decide whether to enroll in college and which type of college and major they wish to enroll in. Let $k \in \mathcal{K}$ denote the field of study and $l \in \mathcal{L}$ denote the type of degree (2-3 year vs 4-year). Let $j \in \mathcal{J}$ denote the decision node in the educational model and $s \in \mathcal{S}$ denote the final schooling level (high school, college dropout or college graduate).

High school graduates begin by choosing whether or not to enroll in college ($D_1$). If they do not enroll ($D_1 = 0$), they enter the high school labor market and earn $Y_1$. If they enroll in college ($D_1 = 1$), they must decide the type of degree and field of study ($D_2(\mathcal{K}, \mathcal{L})$). Once they have entered a college of type $l$ and field $k$, they must decide if they want to change the degree type and/or field of study ($D_{3kl} \in \{0, 1\}$). If they decide to change their initial choice ($D_{3kl} = 1$), then they choose a new degree type and field of study ($D_4(\mathcal{K}, \mathcal{L})$). Finally, they must decide whether to graduate or not ($D_{5,k,l'} \in \{0, 1\}$, where $(k', l')$ denotes the final choice). If they do not graduate they enter the labor market for college drop outs and earn $Y_{2,l'}$, otherwise they enter the labor market for college graduates and earn $Y_{3,k',l'}$. 
4.1 The Labor Market

Associated with each final state \((skl)\) is a potential earnings model for each individual. Let \(Y_{skl}\) denote the earnings of the individual. Earnings are a function of a vector of observables \(X\), a finite dimensional vector of unobserved abilities \(\theta\), and an idiosyncratic error term \(\eta_{skl}\), which is unobserved by the econometrician. We assume a separable model for earnings:

\[
Y_{skl} = \beta_{skl}^X X + \alpha_{skl}^Y \theta + \eta_{skl}
\]

where individual \(i\) subscripts are suppressed.

4.2 Sequential College and Major Choice Model

The decision to enroll \((D_1)\) is characterized by an index threshold-crossing property:

\[
D_1 = \begin{cases} 
1 & \text{if } I_1 \geq 0 \\
0 & \text{otherwise} 
\end{cases}
\]  

(1)
where $I_1$ is the agent’s *perceived* value at node 1 of enrolling in college.

The decision process of degree type and major is characterized by the maximization of a latent variable $I_{jkl}$. Let $I_{jkl}$ represent the perceived value associated with the choice of degree type $l$ and field $k$ at nodes $j \in \{2, 4\}$:

$$D_j(\mathcal{K}, \mathcal{L}) = \arg \max_{l \in \mathcal{L}, k \in \mathcal{K}} \{I_{jkl}\} \text{ for } j \in \{2, 4\}$$

where $D_j(\mathcal{K}, \mathcal{L})$ denotes the individual’s chosen schooling level and field at nodes $j \in \{2, 4\}$.

The decisions to change degree type/major ($D_{3kl}$) and graduate ($D_{5kl}$) are also characterized by an index threshold-crossing property:

$$D_{jkl} = \begin{cases} 1 & \text{if } I_{jkl} \geq 0 \\ 0 & \text{otherwise} \end{cases} \text{ for } j \in \{3, 5\}; \ l \in \mathcal{L}; \ k \in \mathcal{K} \quad (2)$$

where $I_{jkl}$ is the agent’s perceived value of changing degree type/major ($j = 3$) or graduating from college ($j = 5$).

The perceived value(s) for each choice are a function of observable background characteristics ($X$), a finite dimensional vector of unobserved abilities $\theta$, and an idiosyncratic error term $\varepsilon_{jkl}$, which is unobserved by the econometrician:

$$I_{jkl} = \beta_{jkl}^E X + \alpha_{jkl}^E \theta + \varepsilon_{jkl}$$

### 4.3 Measurement System of Unobserved Abilities

Central to our empirical strategy is the existance of a finite dimensional vector ($\theta$) of unobserved endowmenets that generate all of the dependence across the outcomes conditional on the observables $X$. We cannot observe $\theta$, but instead link them to a number of proxies for each dimension of ability. Our estimation strategy accounts for both biases and measurement error in these proxies.
We posit the existence of three underlying latent abilities: cognitive, interpersonal, and grit. Let $M$ denote a vector of measures that define the measurement system for these abilities. We use measures of cognitive and interpersonal abilities from the compulsory Swedish military enlistment to identify cognitive and interpersonal skills. We adjoin data from 9th grade registers on GPA and the choices to take advanced tracks in English and Math. The 9th grade measures are used to identify all three dimensions of ability.

We define $\tilde{M}_n$ as latent variables that map into observed measures $M_n$:

$$M_n = \begin{cases} 
\tilde{M}_n & \text{if } M_n \text{ is continuous} \\
1(\tilde{M}_n \geq 0) & \text{if } M_n \text{ is a binary outcome}
\end{cases} \quad (3)$$

where the latent variables are assumed to be separable in observables, latent abilities and an idiosyncratic error term:

$$\tilde{M}_n = \beta_n^M X + \alpha_n^M \theta + u_n.$$ 

We define a triangular measurement system that describes how each of the abilities load onto the different measures in Table 1.
Table 1: Structure of Measurement System of Abilities

<table>
<thead>
<tr>
<th>Measures</th>
<th>Cognitive</th>
<th>Interpersonal</th>
<th>Grit</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Military Enlistment</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive 1: Inductive</td>
<td>x</td>
<td>b</td>
<td></td>
</tr>
<tr>
<td>Cognitive 2: Verbal</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive 3: Spatial</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive 4: Technical</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leadership Evaluation(^a)</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leadership Ability</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Emotional Stability</td>
<td>x</td>
<td>x(^b)</td>
<td></td>
</tr>
<tr>
<td><strong>9th Grade Education Registers</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9th Grade Advanced Math Track(^a)</td>
<td>x</td>
<td>x</td>
<td>x(^b)</td>
</tr>
<tr>
<td>9th Grade Advanced English Track(^a)</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>9th Grade GPA</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

*Notes: a* Binary discrete choice models. *b* We normalize the loadings in these models so that each factor has standard deviation one.

5 Empirical Result

This section lays out a series of results from the estimated model. In the first sub-section we document sorting into enrollment into college majors, how people reallocate into the their final degrees, and how these both depend on the three skills used in the model. In the second sub-section we document a range of treatment effects across the eight 4-year majors we consider where the treatment effects are calculated as the returns to these specific majors over not enrolling in college. These treatment effects are broken down into the effect of enrolling in a given major (given people may then reallocate or drop out) and the effect of earning a particular major. In the third sub-section we consider a set of policy relevant treatment effects. These treatment effects are the returns to a particular major for individuals who did not enroll in major but were almost indifferent between enrolling in their chosen major and the major under consideration. For this analysis, the treatment effects are calculated as the gains of choosing a particular major over the major they would have otherwise selected (for the population close to indifferent). The last two sub-sections lay out marginal treatment effects for enrollment. The first shows the returns from enrolling based on deciles of the three skills. The second shows the
returns from enrolling as a function of the probability that major was selected. As this section presents many results, it provides some results for the eight majors being studied, but at some places will focus on Humanities, Science and math, and Engineering.

5.1 Sorting into College Majors

This subsection provides additional details on how individuals reallocate across majors while in school and how individuals sort into first and final major based on skill. Table 2 shows how people that enroll in college sort into initial major and then reallocate into the major with which they graduate. The rows represent the major individuals initially enroll in and the columns represent final majors. The first panel shows the proportion ending up in each final major based on their initial enrollment while the second panel shows the proportion in each initial enrollment based on their final degree. For example, the table shows that 45% of people who enroll in business degrees do not complete the degree, with most not completing any 4-year degree. In contrast, 79% of those that enroll in medicine finish their degree, with some reallocating so social science or non-4-year majors.

Figure 2 shows how students sort into college major enrollment and graduation by skill. The top panel shows the average levels of the three skills based on initial college enrollment while the bottom panel shows the average levels of the three skills based on major upon graduation. For both panels all three skills have been normalized to be mean 0 standard deviation 1 for the population of people who ever enroll in college. The figure shows that those who enroll in humanities degrees tend to be below average in all three skills. In contrast, those who enroll in medicine tend to be above average in all three skills. For other majors there is differential sorting on skill. For example, business majors tend to be high in grit and inter-personal skill but below average in cognitive skill.Math and science majors tend to above average in cognitive skill but below average in grit and inter-personal skills.
Figure 2: Sorting into Major Enrollment and Graduation By Skill

Enrollment

Graduation

Notes: Table shows the average inter-personal, cognitive, and grit skills by major. All skills are normalized to be mean 0 and standard deviation one for the population of people who ever enroll in college. The top panel shows average skills by initial enrollment and the bottom panel shows average skills by obtained major.
Table 2: Enrollment and Completion of College Majors

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Business</td>
<td>0.45</td>
<td>0.01</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
<td>0.00</td>
<td>0.02</td>
<td>0.03</td>
<td>0.46</td>
</tr>
<tr>
<td>Education</td>
<td>0.00</td>
<td>0.55</td>
<td>0.01</td>
<td>0.03</td>
<td>0.00</td>
<td>0.00</td>
<td>0.03</td>
<td>0.01</td>
<td>0.37</td>
</tr>
<tr>
<td>Engineering</td>
<td>0.01</td>
<td>0.00</td>
<td>0.64</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.02</td>
<td>0.01</td>
<td>0.31</td>
</tr>
<tr>
<td>Humanities</td>
<td>0.00</td>
<td>0.03</td>
<td>0.01</td>
<td>0.27</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
<td>0.05</td>
<td>0.63</td>
</tr>
<tr>
<td>Law</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
<td>0.00</td>
<td>0.63</td>
<td>0.01</td>
<td>0.01</td>
<td>0.02</td>
<td>0.30</td>
</tr>
<tr>
<td>Medicine</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
<td>0.01</td>
<td>0.00</td>
<td>0.79</td>
<td>0.01</td>
<td>0.06</td>
<td>0.11</td>
</tr>
<tr>
<td>Science and Math</td>
<td>0.01</td>
<td>0.02</td>
<td>0.04</td>
<td>0.00</td>
<td>0.00</td>
<td>0.02</td>
<td>0.36</td>
<td>0.02</td>
<td>0.53</td>
</tr>
<tr>
<td>Social Sciences</td>
<td>0.05</td>
<td>0.01</td>
<td>0.03</td>
<td>0.01</td>
<td>0.03</td>
<td>0.01</td>
<td>0.02</td>
<td>0.24</td>
<td>0.59</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Business</td>
<td>0.87</td>
<td>0.02</td>
<td>0.01</td>
<td>0.02</td>
<td>0.06</td>
<td>0.00</td>
<td>0.05</td>
<td>0.12</td>
<td>0.18</td>
</tr>
<tr>
<td>Education</td>
<td>0.01</td>
<td>0.92</td>
<td>0.00</td>
<td>0.36</td>
<td>0.02</td>
<td>0.01</td>
<td>0.07</td>
<td>0.07</td>
<td>0.16</td>
</tr>
<tr>
<td>Engineering</td>
<td>0.04</td>
<td>0.02</td>
<td>0.95</td>
<td>0.10</td>
<td>0.04</td>
<td>0.12</td>
<td>0.08</td>
<td>0.07</td>
<td>0.28</td>
</tr>
<tr>
<td>Humanities</td>
<td>0.00</td>
<td>0.01</td>
<td>0.00</td>
<td>0.38</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.02</td>
<td>0.03</td>
</tr>
<tr>
<td>Law</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.81</td>
<td>0.01</td>
<td>0.00</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>Medicine</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.72</td>
<td>0.00</td>
<td>0.04</td>
<td>0.04</td>
<td>0.01</td>
</tr>
<tr>
<td>Science and Math</td>
<td>0.01</td>
<td>0.02</td>
<td>0.03</td>
<td>0.04</td>
<td>0.02</td>
<td>0.08</td>
<td>0.77</td>
<td>0.07</td>
<td>0.19</td>
</tr>
<tr>
<td>Social Sciences</td>
<td>0.06</td>
<td>0.01</td>
<td>0.01</td>
<td>0.07</td>
<td>0.06</td>
<td>0.04</td>
<td>0.03</td>
<td>0.57</td>
<td>0.12</td>
</tr>
</tbody>
</table>

Notes: Table shows transition matrix from majors enrolled in (rows) and final degree attainment (columns) for those who initially enrolled in one of the eight listed majors. The first panel shows the proportion ending up in each degree by enrollment. The second panel shows the proportion enrolling in each degree by final degree attainment. The third panel shows the number of observations in each enrollment / attainment cell.

5.2 Treatment Effects

This section estimates the treatment effects for the eight 4-year majors. Specifically, we estimate the gains from enrolling in or obtaining a degree in a particular major over being a high school graduate with no college. The estimates are for the population of individuals who ever enroll in college and are estimated both for enrolling in the major (where individuals may reallocate) and graduating with a four year degree in that particular major. For each, we calculate the average treatment effect, the average treatment effect for those with high skill levels, the average treatment effect for those with low skill levels, the treatment on the treated (TT), and the treatment on the untreated (TUT).\footnote{High skill is defined as being in the top 50\% of all three skills while low skill is defined as being in the bottom 50\% of all three skills.}
Figure 3 shows the full set of treatment effects for the eight four year majors. The top panel shows the treatment effects from enrollment and the bottom panel shows the treatment effects from graduation. The Figure shows that the average returns to most college majors are positive, with the exception of Humanities, which has a negative return over being a high school graduate. Except for education we see that the expected returns to enrollment are higher for those who enroll (treatment on the treated) compared to the expected returns to those who don’t enroll in the major (treatment on the untreated). Selection on gains is particularly strong for Medicine and law.

Figure 3 also shows that there are differential returns to skill across various majors. For example, the returns to a business or law degree are much higher for high ability individuals than for low ability individuals. In contrast, the returns to majors such as Social Science vary less by skill.

Finally, Figure 3 can tell us about the relative returns to enrollment in a major compared to completing a major. This is particularly important as many majors have low completion rates and one of the driving forces behind low-skill individuals pursuing sum degrees may be the difficulty of completing the degree rather than the returns the degree would provide. For example, there are large differences in TT versus TUT and ATE(high) versus ATE(low) for enrolling in Medicine, but the estimated returns to graduation vary less by skill and selection.
Figure 3: Treatment Effects by College Major (enrollment and graduation)

Enrollment
Treatment Effects: Returns to Enrolling by Major

Graduation
Treatment Effects: Returns to Graduation by Major

Notes: Table shows the estimated treatment effects for enrolling (top panel) and graduating (bottom panel) from particular college majors. The treatment effects are estimated for everyone who ever enrolls in high school. High ability is defined as being in the top half of all three skill distributions while low ability is defined as being in the bottom half of all three skill distributions.
5.3 Policy Relevant Treatment Effects

The previous section estimated the returns to majors over being a high school graduate and for the full population of those who enroll in college. While interesting, these treatment effects do not correspond to any specific policy which could be implemented. This section constructs policy relevant treatment effects, by calculating the returns to a particular major for the population of people who almost chose to enroll in that major but did not. The treatment effect is then calculated as the difference in their earnings between their earnings in that major and the major they actually chose. In this way, the treatment effects correspond to the benefits for individuals who would be induced into a given major by making that major slightly more attractive, and the gains are estimated against what they would have otherwise pursued.

Figure 4 shows the average gains to each major for those who did not select the major but were almost indifferent. This is the policy-relevant population as these are individuals who would have selected the major under policies that uniformly made the major marginally more appealing. In addition, the comparison treatment effects are estimated as the returns to choosing a particular major versus choosing the major they would have otherwise chosen. Across the majors we see that the are large differences in the PRTE. For humanities or education, the PRTE for enrollment is negative. In contrast, the PRTE for Engineering, Medicine, Business, and Law are large and positive.

Figure 4 tells us about the average PRTE for each major, but does not tell us how this depends on which alternative majors individuals are leaving behind or how the PRTE differs across these groups. Figures 5, 6, and 7 show the PRTE for Humanities, Medicine, and Business broken down into which majors they are induced out of. The top panels shows the PRTE for those induced to leave the major listed along the x-axis, while the bottom panel shows what proportion of those nearly indifferent came from each of the 12 majors. For example, for Humanities, many individuals left 3-year Non-STEM degrees resulting in a small loss of earnings. In contrast, for Medicine, most are induced in from Engineering, and these individuals have large returns.
Figure 4: Policy Relevant Treatment Effects by Major (enrollment)

Notes: Table Shows the average treatment effect for those who chose other majors but were almost indifferent to that major and the major listed along the x-axis. The treatment effects are the impact of switching to the major along the x-axis versus remaining in their original choice.
Figure 5: PRTE: Humanities

Notes: Table Shows the average treatment for those just indifferent to entering the major. The Treatment effect is returns to the major minus the returns to what was the agent’s first choice. The panel on the left shows the returns for those nearly indifferent based on which major they are induced to leave. The right panel shows what proportion of those induced into the major were in each of the other majors.
Figure 6: PRTE: Medicine

Notes: Table shows the average treatment for those just indifferent to entering the major. The Treatment effect is returns to the major minus the returns to what was the agent’s first choice. The panel on the left shows the returns for those nearly indifferent based on which major they are induced to leave. The right panel shows what proportion of those induced into the major were in each of the other majors.
Figure 7: PRTE: Business

Notes: Table Shows the average treatment for those just indifferent to entering the major. The Treatment effect is returns to the major minus the returns to what was the agent’s first choice. The panel on the left shows the returns for those nearly indifferent based on which major they are induced to leave. The right panel shows what proportion of those induced into the major were in each of the other majors.
5.4 Returns to Major By Skill

This section reports the average returns to each major by decile of the three skills. The treatment effects are the benefits of enrolling or graduating with that degree over being a high school graduate. Each plot below shows the expected treatment effect given being in that decile of the particular skill (all other covariates and skills integrated out).

Figures 8, 7, and 6 show the returns by skill for Humanities, Business, and Medicine degrees. The top panel shows the returns for each skill for enrollment while the bottom panel shows the returns to each skill for graduation.
Figure 8: Returns by Skill: Humanities

Notes: Figures show the marginal return to skill for particular college major. Top panel shows the benefits of enrolling while the bottom panel shows the benefits of graduating.
Figure 9: Returns by Skill: Medicine

Notes: Figures show the marginal return to skill for particular college major. Top panel shows the benefits of enrolling while the bottom panel shows the benefits of graduating.
Figure 10: Returns by Skill: Business

Notes: Figures show the marginal return to skill for particular college major. Top panel shows the benefits of enrolling while the bottom panel shows the benefits of graduating.
5.5 Marginal Treatment Effects

This section reports the average treatment effects to enrolling in particular majors as a function of the probability that individuals choose to enroll in that major. This allows us to see if there is selection on gains across majors. As in the previous section, the treatment effect is the returns of enrolling in that particular major over the being a high school graduate.

Figures 11, 12, and 13 show the marginal treatment effects. For all three of these majors there is evidence of selection on gains, with the returns being higher for those more likely to select the major. Selection on gains is particularly pronounced for enrollment in science and math majors.$^{13}$

$^{13}$Note that the table only show point estimates for percentiles of the probability that have at least 500 observations using a simulation of ten million agents.
Figure 11: MTE: Science and Math

Notes: Figures show the average returns of enrolling the specific major over the returns to being a high school graduate as a function of the probability the major was initially chosen for enrollment.

Figure 12: MTE: Engineering

Notes: Figures show the average returns of enrolling the specific major over the returns to being a high school graduate as a function of the probability the major was initially chosen for enrollment.
Notes: Figures show the average returns of enrolling the specific major over the returns to being a high school graduate as a function of the probability the major was initially chosen for enrollment.

References


