Major Complexity Index and College Skill Production *†‡

Xiaoxiao Li

Sebastian Linde

Hajime Shimao

July, 2023

Abstract

We propose an easily computable measure called the Major Complexity Index (MCI) that captures the latent skills imparted by various college majors. Specifically, we apply the Method of Reflections, which is an iterative algorithm originated in the context of international trade, to parse through a major-to-occupation flow network and formulate a scalar measure that encapsulates the relative skill complexity of majors that is distinct from existing metrics pertaining to major specificity. Our complexity measure appears to be a potent factor in explaining individual earning and employment differences across college majors, and the results remain robust to confounding factors and aggregation issues. Further results suggest that the MCI can not only account for current income disparities but also predict future major-level earning growth, a feat beyond the capability of typical major specificity indices. Additional exercises reveal that the MCI strongly relates to advanced skills such as critical and analytical thinking, as well as abilities to analyze and solve quantitative, practical, and complex problems.

Keywords: Return to College Major; Major-to-Occupation Network; Skill Production.
JEL classifications: I2, J2

*We are deeply indebted to Timothy Bond, Caroline Hoxby, Michael Lovenheim, and Kevin Stange, and two anonymous referees for suggestions and advice that helped us improve the manuscript substantially. We are also grateful to Laura Ahlstrom, Joseph Altonji, Francisca Antman, Julian Betts, Barbara Biasi, Peter Q. Blair, Celeste Carruthers, Aimee Chin, Weili Ding, Jane Fruehwirth, Clint Harris, Teresa Harrison, David Hummels, Christopher Kilby, Elif Kubilay, Laura Meinzen-Dick, Lois Miller, Kevin Mumford, Philip Oreopoulos, Steve Ross, Jeffrey Smith, Yang Song, Jones Todd, Evan Totty, Sarah Turner, Muzhe Yang, and participants at EEA, WEAI, 2021 CSWEP CeMENT Mentoring Workshop, MEA, NBER Economics of Education Spring 2021 Program Meeting, and SOLE, ASSA 2023, Swarthmore Brown Bag, for helpful comments. We would like to extend our sincere gratitude to Margaret Leighton and Jamin Speer for generously providing their code for the computation of the Gini coefficient, and appreciation to Audrey Light and Sydney Schreiner Wertz for their invaluable assistance in clarifying our understanding of major construction in the NLSY97.

†The authors declare no relevant or material financial interests that relate to the research described in this paper.

‡Contact information: Xiaoxiao Li (corresponding author), Villanova University, xiaoxiao.s.li@villanova.edu, (610) 519-3256; Sebastian Linde, Medical College of Wisconsin, slinde@mcw.edu; Hajime Shimao, Santa Fe Institute and Pennsylvania State University, hjs5825@psu.edu.

1 Introduction

The return to education varies widely across fields of study in college (Altonji et al., 2012, 2016; Lovenheim and Smith, 2023). While some of this disparity is certainly attributed to the selection process (Arcidiacono, 2004; Webber, 2014), a significant portion of it stems from the different skillsets acquired in various academic programs (Kirkeboen et al., 2016; Hastings et al., 2013; Daly et al., 2022; Andrews et al., 2017; Bleemer and Mehta, 2022). In other words, not all majors are created equal when it comes to equipping students with marketable skills. As Hemelt et al. (2021) aptly point out, each college major can be viewed as a "portable bundle of skills", especially from employers' perspectives. Thus, understanding the nature and implications of these skill bundles is key to unraveling the puzzle of major-dependent return on education.

The diversity in skills acquired across academic disciplines naturally leads to varying occupational destinations for graduates, and can potentially drive the differential returns to college majors (Lovenheim and Smith, 2023). For instance, a petroleum engineering degree can pave the way to lucrative careers as petroleum engineers, an opportunity often not as readily accessible to students from other fields of study. This intersection between college majors and subsequent occupational outcomes contains rich information to be exploited, and can be seen as a testament to the matching of multifaceted skills.

In this paper, we extend recent development in the space of quantifying major skill bundles and specificity (Altonji et al., 2012; Blom et al., 2021; Leighton and Speer, 2020; Hemelt et al., 2021) by proposing a novel index derived from the major-to-occupation network. Specifically, we utilize the "Method of Reflections" technique, first introduced by Hidalgo and Hausmann (2009) in the context of international trade and economic growth[1] , and develop a unique index called the Major Complexity Index (MCI) which harnesses the network structure between majors and occupations.

[1]This method was initially designed to examine the bipartite network between countries and their exported goods, aiming to quantify the latent production capacities and technologies that enable a country to produce its observed basket of exports. Following its introduction, this technique and the resultant metrics of economic complexity (ECI) have been extensively utilized to scrutinize the influence of economic structures on outcomes including economic growth, income inequality, greenhouse emissions, employment, and the spatial concentration of economic activities. See Hidalgo, 2021; Balland et al., 2022 for a detailed summary of its applications.

The Method of Reflections capitalizes on the richness of information contained within the major-to-occupation network, extracting the latent structure of necessary skill sets through an iterative algorithm. The fundamental idea is that occupations requiring a complex skill set tend to be associated with majors that cultivate such skills, and likewise, majors that equip students with more complex skills often lead them to occupations that demand a higher degree of proficiency. According to this paradigm, the complexity of majors and occupations reciprocally define each other, and this intrinsic link demands a recursive construction of the complexity index, reflecting the intricate interplay between majors, skills, and occupational demands.

To our knowledge, this study represents the first application of the Method of Reflections within the realms of education and labor markets, heralding a novel approach to quantify the skill bundles of college majors. In deference to the original complexity economics literature initiated by Hidalgo and Hausmann (2009), we refer to this index as major "complexity" of skills, as it encapsulates the intricacies and nuances of thousands of teaching and learning activities embedded in the knowledge and skill production process, as well as the sophisticated ways these building blocks combine to form the holistic skill sets of individuals.

In contrast to existing major specificity indexes, such as the occupational Herfindahl-Hirschman Index (HHI) which summarizes the occupation concentration without accounting for the type of occupations associated with each major (Blom et al., 2021), our approach utilizes the information embedded in an entire bipartite major-to-occupation matrix to reveal an underlying tripartite structure, connecting college majors to the skills they cultivate, and occupations to the skills they demand. As a result, the MCI effectively captures the relevant neighboring information within the network, allowing majors with the same degree of specificity to have markedly different complexity rankings. Consider the example of Petroleum Engineering and Mathematics Teacher Education. Both these majors are highly specialized according to the HHI. However, their MCI rankings diverge significantly. Petroleum Engineering, which is associated with occupations rarely accessible to majors of low complexity, and consequentially is connected with other complex majors, yielding a high MCI rank (14/159). In contrast, Mathematics Teacher Education, while also specialized, leads students to occupations that are commonly accessible and thus is connected to other low-complexity majors, resulting in a lower MCI rank (125/159).

Our empirical investigation mainly employs individual-level data from the American Community Survey (ACS), spanning from 2009 to 2019. This large pooled dataset enables us to construct a representative bipartite major-to-occupation network, which serves as the foundation for our Major Complexity Index. Through extensive analyses, we investigate the relationship between the MCI and labor market outcomes across individuals including earnings and employment differentials. Our findings provide compelling evidence that the MCI captures significant aspects of college majors that influence both outcomes for graduates. Specifically, a one standard deviation increase in the MCI corresponds to an 8% increase in salary and a 1.56% boost in employment.

An intriguing finding of our study is that the explanatory power of the MCI is nearly orthogonal to that of existing metrics pertaining to major specificity such as the HHI, indicating that the MCI is not merely a recapitulation of existing measures. This is in a way a very surprising outcome, considering that the data utilized to compute the specificity indices, like the HHI, and the complexity index are entirely identical. Both indices employ the distributional differences in occupational outcomes across various majors to infer certain aspects of skill acquisition. Nevertheless, our results indicate that the Method of Reflections successfully extracts information from a major-to-occupation network that is distinct and complementary to what is measured by existing methods. This finding paves a new way of utilizing occupational placements of college graduates to understand the variations in skill acquisition across different majors.

It is worth noting that, even within the same major categories (e.g., Science, Business, etc.) and occupational groups (SOC2), majors with higher complexity still yield higher income returns. Additional analyses demonstrate the robustness of the MCI in explaining labor market outcomes, in relation to aggregation issues and confounding factors. Lastly, our heterogeneity analysis indicates that the complexity of majors has lasting impacts on career outcomes, even though the returns may not be evenly distributed across different demographic groups.

Furthermore, drawing parallels to the work of Hidalgo and Hausmann (2009), who demonstrated a robust correlation between the Economic Complexity Index (derived from country-to-exported-product networks) and a country's GDP growth, we perform a prediction of future return to college majors using the MCI. Our findings uncover a significant association between the annually constructed MCI and future earnings linked to different majors. Specifically, a one standard deviation increment in the current MCI is correlated with a salary increase of 2.08% in five years' time, and 2.84% in ten years' time. In contrast, the HHI, while highly correlated with current income level, fails to demonstrate a comparable predictive power for future changes. This differentiation further accentuates the unique insights the MCI offers in understanding both present and future earning dynamics across academic majors.

To deepen our understanding of what the MCI measures, we merged the MCI computed from the ACS with major-level characteristics derived from the National Survey of Student Engagement (NSSE). The analysis reveals that high-MCI majors tend to attract students with superior pre-college academic qualifications, as evidenced by higher SAT scores. Furthermore, high-MCI majors appear to be more intensive and demanding as students in these majors typically devote more time preparing for classes, completing problem sets, and working on longer written assignments. Interestingly, when it comes to the knowledge and skills acquired through college education, students in high-MCI majors frequently report marked development in thinking critically and analytically, as well as enhanced abilities to analyze and solve quantitative, practical, and complex problems. However, there seems to be no correlation between the MCI and basic skills such as written and spoken communication, which are presumably developed mainly through primary and secondary education. It is also worth noting that, the MCI exhibits a negative correlation with broad general education, and no correlation with job-specific knowledge and skills, which suggests that the MCI transcends the traditional distinction between general and job-specific skills.

Though the MCI correlates with measures of student ability, its power in explaining labor market outcomes remains robust and consistent even after controlling for a range of individual characteristics. This includes not only markers of student ability, but also factors pertaining to family background. We substantiate this finding by integrating our ACS-derived MCI values with data from the National Longitudinal Survey of Youth 1997 (NLSY97). These findings highlight the MCI's potential as a valuable tool for comprehending the complexity of the links between college majors, skills cultivation, and labor market success.

The rest of the paper is organized as follows: Section 2 summarizes the literature and highlights our connection to the most relevant body of works. Section 3 introduces the Method of Reflections in the context of a major-to-occupation network. Section 4 details the data sources. Section 5 presents our empirical results, and Section 6 discusses important policy implications, limitations, and possible future extensions. Supplementary materials are included in Appendices A-G.

2 Literature

To comprehend the variations in skills taught across different majors and examine the subsequent labor market consequences, prior research has focused on the specificity of majors concerning their resulting occupational placements, since there can exist trade-offs between specialized skills that target a set of occupations and general skills that, though less productive, can easily be transferred across occupations (Silos and Smith, 2015). For instance, Altonji et al. (2012); Martin (2022) utilize the proportion of students from each major that ended up in the top three occupations (Top3). Pursuing a more refined approach, Blom et al. (2021) compute the Herfindahl-Hirschman Index (HHI) to measure the concentration of occupational placements for each major. These studies highlight that majors with higher specificity, that is, those leading to a concentrated range of specific occupations, yield a higher return on investment. Kinsler and Pavan (2015) investigates the monetary returns of three broadly defined college majors. Utilizing a survey question of the relatedness of one's job to the undergraduate field of study, they implement a dynamic structural Roy model with two types of human capital and find that individuals tend to earn more when their occupation is closely aligned with their college major. Recently, Leighton and Speer (2020) proposed an alternate measure of major specificity utilizing the Gini coefficient that represents the distribution of returns to majors across various occupations. Their findings suggest that majors with higher Gini coefficients, indicating uneven earnings across occupations, are more specific and experience higher labor market returns.

While these studies underscore the importance of a major's skill specificity, they overlook the nature of the occupations into which these majors channel their students. For instance, consider two very specialized majors: Computer Engineering and Elementary Education. Both of these majors exhibit a high degree of specificity, leading to high scores in the Top3 share, HHI, and Gini coefficients. However, the occupations they funnel their graduates into are very different. A vast majority of Computer Engineering graduates work as computer engineers, while Elementary Education students predominantly become elementary school teachers. The specificity indices employed in these studies fail to capture such critical distinctions between the occupations that different majors feed into. This highlights a need for more nuanced measures that consider not just the occupation concentration or wage dispersion, but also the nature of the occupations related to each major. However, though it is intuitively clear that occupations such as computer engineers and elementary school teachers are distinctly different, articulating these differences can pose a significant challenge. In this paper, we introduce a novel approach to tackle this issue by exploiting the rich information embedded in a major-to-occupation flow network. Specifically, we characterize occupations based on their associations with different majors, while the majors are characterized by the occupations they are linked to.

Overall, our work contributes to a rich literature on skill formation in college. This literature has been confronted with important challenges that we believe our approach is able to circumvent. Firstly, knowledge and skills produced in majors are unobservable and extremely difficult to quantify, particularly along the dimensions that are of interest to employers (with Hemelt et al. 2021 being a notable exception, as detailed below). For instance, how do we incisively measure the programming skills that students can obtain from an economics major on average?

Secondly, the skills deemed significant within the labor market are presumably of high dimensionality, as explored by Hemelt et al. (2021). Indices specific to academic majors, such as the HHI or the MCI, can delve into the nature of these skills without necessitating an explicit model of their dimensionality. Meanwhile, in instances where these skills are explicitly modeled, they are typically defined within relatively low dimensions, predominantly based on intuitive understandings. As an illustrative example, Cunha and Heckman (2007) identify two principal dimensions of skills—cognitive and non-cognitive—via a factor model in their seminal work.[2] For reasons of computational feasibility, Kinsler and Pavan (2015) focus on only two particular types of human capital, namely mathematical and verbal. While these dimensions may indeed represent critical distinctions in skills, adopting a narrow focus on them could be misleading within the context of post-secondary education and occupational outcomes. For instance, leadership—a specific skill within the non-cognitive domain—has been documented to have predictive capacity in terms of potential earnings (Kuhn and Weinberger, 2005). Recent work by Deming (2017) underscores the growing importance of social skills within the labor market.

Similarly, another important yet under-explored aspect is the complementarity among skills (Cunha et al., 2006). A job task often requires a combination of skills. For example, to be a financial engineer, one not only needs to be skilled in financial econometrics and programming but also in management and communication that complement the technical background and help to improve job performance. Even if we fully observe the skill production process, with high dimensionality, it rapidly becomes impossible to estimate the complementary effects of every combination of fine-grained skill categories. For instance, Hemelt et al. (2021) exploit job vacancy data that is free from occupational sorting and can describe majors by a vector of 11 exclusive skill composite categories and 1,000 most frequently listed skills. However, with finite data variations, it is difficult to fully explore the complementary effects among the detailed skill categories.

In brief, the challenges described above make our proposed methodology particularly appealing. Rather than striving to resolve these issues, our approach seeks to deftly circumvent them. The "building-block" model, consisting of a tripartite network of majors-skills-occupations, encapsulates the high-dimensional and combinatorial nature of skills. As theoretically demonstrated in Hausmann and Hidalgo (2011), the iterative procedure of the Method of Reflection can uncover the underlying tripartite structure hidden within the bipartite network. Consequently, the MCI

[2] Note, the factor analysis approach of Cunha and Heckman (2007, and subsequent papers) can, in principle, accommodate a broader dimension of skills.

preserves the crucial structural aspects of skill matching, while its computation sidesteps the need to explicitly model the entirety of the skill structure. Instead, the calculation of the MCI directly reveals the relative skill complexity value of college majors. It achieves this by exploiting the match between students from majors and subsequent occupational placement. In effect, this methodology allows us to draw insights from the complexity of skills without becoming mired in their intricate dimensions—a key advantage in navigating the landscape of education and labor market research.

There are surprisingly few easily-computable quantitative descriptions of college majors, with only a handful of occupation-based major specificity measures such as Top3 share (Altonji et al., 2012), HHI (Blom et al., 2021), and Gini coefficient (Leighton and Speer, 2020) being among the available tools.[3] Surely, the characterization of majors goes beyond the conventional delineation between general and specific skills. For instance, Hemelt et al. (2021) leverage job ads data to illustrate majors via a detailed skills demand vector. They reveal significant variations in both skill demand and earnings across various majors and regions, leading to the conclusion that majors can be viewed as portable skill bundles. Nonetheless, such data is not typically easily accessible.

Our comprehensive Major Complexity Index (MCI) significantly contributes to this landscape by offering a simple-to-compute measure that requires minimal data. Crucially, it doesn't rely on student-reported data on major features as used in Kinsler and Pavan (2015). This method builds on recent advancements in this field by exploiting the rich information hidden in the major-to-occupation network, thereby complementing existing structure-based approaches (see, for example, Altonji et al. 2012; Kinsler and Pavan 2015). In this manner, it enhances our understanding of the otherwise unseen skill production process associated with college majors. Moreover, the MCI can serve as an informative and convenient reference for a range of stakeholders. Prospective students can use it as a tool in choosing their college majors, and education administrators can draw upon it in strategic planning, particularly in resource-limited environments.

[3] There is also a curriculum-based approach that draws upon transcript data and summarizes major based on credits or GPA across subjects. See, e.g., Hamermesh and Donald (2008); Silos and Smith (2015); Light and Schreiner (2019).

3 Method

Suppose we have $\mathscr{M}$ college majors and $\mathscr{J}$ occupations. Let M be a $\mathscr{M} \times \mathscr{J}$ flow matrix that represents the majors to occupations network. We consider the flow matrix to be a binary matrix where $M_{m,j} = 1$ if major m places a significant proportion of students in occupation j, and $M_{m,j} = 0$ otherwise. Using the Method of Reflections introduced by Hidalgo and Hausmann (2009), we iteratively calculate the value for each major and occupation, respectively. To illustrate the method, suppose there are four majors and four occupations, linked in a network as shown in Figure 1.

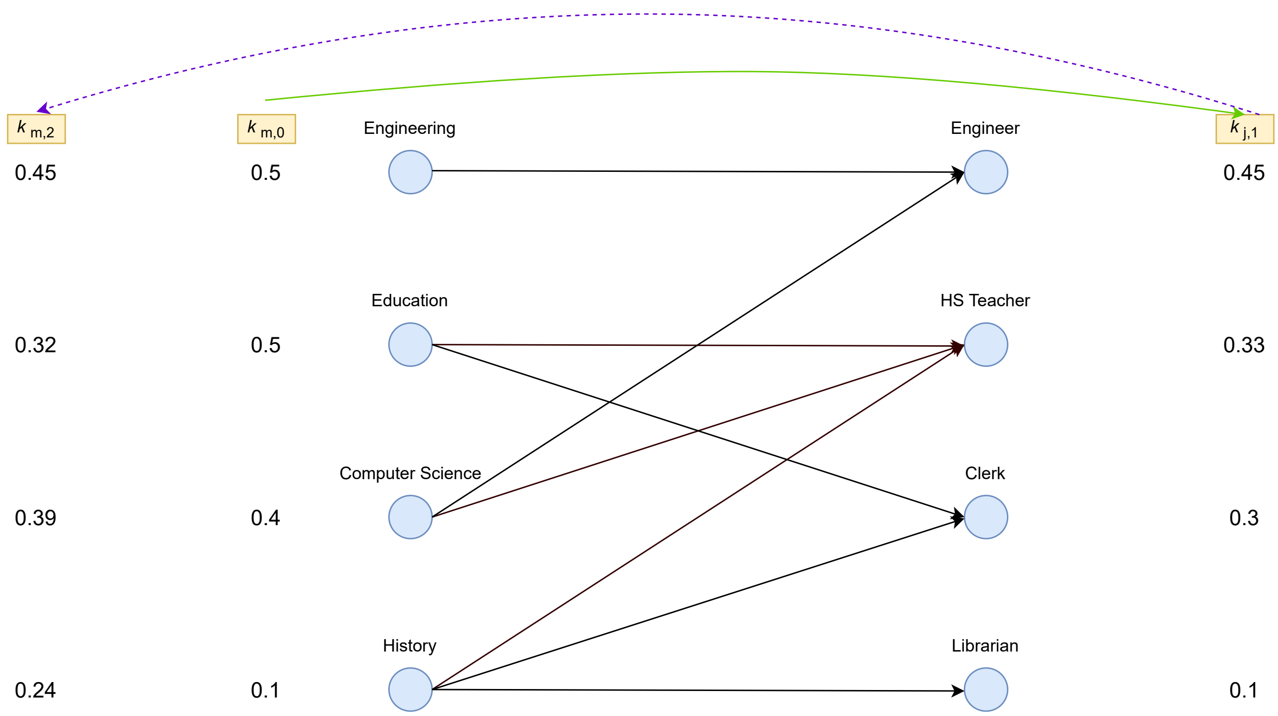

Figure 1: Example

We initiate the process with an initial value, denoted as $k_{m,0}$, that characterizes the skills associated with each major. The particular choice of this initial value has minimal impact on the converged Major Complexity Index (MCI) as we discuss below. As an illustrative starting point, let's consider this initial value to be the occupational Herfindahl-Hirschman Index (HHI), as utilized in Blom et al. (2021), which is recognized for capturing the specificity of majors. Majors that concentrate most students to a small set of occupations are considered specialized, and those that distribute graduates evenly across many occupations are considered general. However, the HHI does not factor in the detailed relationships between individual majors and occupations, thereby not accounting for what type of occupations are associated with each major, and ignores valuable information from other neighboring majors that map into the same occupations in the network. Using the examples in Figure 1, the HHI of Engineering and Education majors are both 0.5. The difference is that Engineering can lead students to become Engineers, which is not easily accessible to students from low-HHI majors like History. Meanwhile, Education sends students to High School (HS) Teachers and Clerks, both accessible from History. Nonetheless, using the HHI directly, Engineering and Education are both ranked in first place, followed by Computer Science, and then History.

To incorporate the accessibility of occupations within the network, we exploit an iterative procedure called the Method of Reflections. We first use the initial index described above, $k_{m,0}$, to calculate the complexity of each occupation, $k_{j,1}$ in the first iteration ($b = 1$), as

$$k_{j,1} = \frac{1}{D_j} \sum_m M_{m,j} \, k_{m,0} \ ,$$

where D_j is the node degree (i.e. the number of major links) of an occupation j, and again $M_{m,j} = 1$ if major m places a significant proportion of students in occupation j, and 0 otherwise. See the solid arrow at the top of Figure 1 for a graphic illustration of this iteration from $k_{m,0}$ to $k_{j,1}$. As an example, for $j = HS\ Teacher$,

$$k_{j,1} = \frac{0 \times 0.5 + 1 \times 0.5 + 1 \times 0.4 + 1 \times 0.1}{3} = \frac{0.5 + 0.4 + 0.1}{3} \approx 0.33.$$

That is, there are three majors, Education, Computer Science, and History that place a significant proportion of students as HS Teachers ($D_j = 3$, $M_{m,j} = 1$ only for those majors), and the average HHI ($k_{m,0}$) of these three majors is calculated to be $0.3\dot{3}$. The same calculation can be carried out for all occupations $j \in \mathscr{J}$, as shown in Figure 1, $k_{j,1}$ column. Using examples in this network, Engineer and Clerk are both connected to two majors, but the HHI of majors that are linked to Engineer is, on average, greater than that for Clerk. As a result, Engineer is ranked higher than

Clerk in this iteration. Importantly, majors are connected to one another through the occupation nodes within a network, and this intermediate step $k_{j,1}$ is essential to reflect such connections.

We then iterate back to the major side ($b = 2$) to update the index of each major to $k_{m,2}$, using those values obtained in $k_{j,1}$ as

$$k_{m,2} = \frac{1}{D_m} \sum_j M_{m,j} \, k_{j,1} \, .$$

where D_m is the node degree (i.e. the number of occupation links) of a major m. See the dashed arrows on the top of Figure 1 for a graphic illustration of this iteration. Using the example of $m = Education$,

$$k_{m,2} = \frac{0.33 + 0.3}{2} \approx 0.32,$$

since $D_m = 2$, and $M_{m,j} = 1$ for $j \in \{HS\ Teacher,\ Clerk\}$ as Education major is only linked to these two occupations. Intuitively, the average score of $k_{j,1}$ for occupations linked to Education is 0.32. Recall, $k_{j,1}$ captures the average initial value (HHI in this case) of majors that are connected to an occupation. Following this logic, $k_{m,2}$ is basically the average of average HHI of majors that are connected to one another through common occupational outcomes. What's embedded in this averaging process is the adjustment of the initial major-level characteristics (e.g., HHI) by occupation accessibility. As shown in Figure 1, the index of Computer Science is updated from $k_{m,0} = 0.4$ to $k_{m,2} = 0.39$, while the adjustment for the Education major is from $k_{m,0} = 0.5$ to $k_{m,2} = 0.32$ in this iteration. As a result, Computer Science is now ranked higher than the Education major, for the reason that Computer Science can send students to occupations, such as Engineers, that are difficult to get into from low-HHI majors, and in turn, it is connected to other majors that are relatively more specific in terms of skills taught on average.

It is important to note that the updated index at the major side is no longer identical to the initial values. This is because the iterative procedure reflects the deeper and more intricate connections in the major-to-occupation network. This iterative process, along with the information passed through the network, leads to the final MCI value that is almost orthogonal to the initial HHI values. In other words, MCI and HHI, though originating from the same data, provide significantly different insights into the complexity of the skills associated with different majors.[4] As we show in the empirical section, the converged MCI starting from the HHI is almost orthogonal to the initial HHI. Furthermore, as we discuss below, the particular choice of initial value has minimal impact on the converged value. Henceforward, in honoring the Complexity Economics literature, we refer to the updated index on the major side as a measure of major complexity, given that it encapsulates the intricacies and nuances of thousands of teaching and learning activities embedded in the knowledge and skill production process, as well as the sophisticated ways these building blocks combine to form the holistic skill sets of individuals (Hidalgo, 2021).

While we stop the illustration at $b = 2$ (where b refers to the number of iterations), we iterate on this procedure for each major and occupation according to equations (1) and (2), respectively, until the major ranking exhibits full convergence:

$$k_{m,b} = \frac{1}{D_m} \sum_{j \in \mathscr{J}} M_{m,j}\, k_{j,b-1} \ , \tag{1}$$

$$k_{j,b} = \frac{1}{D_j} \sum_{m \in \mathscr{M}} M_{m,j}\, k_{m,b-1} \ . \tag{2}$$

The complexity index on the occupation side, $k_{j,b}$ for any odd number b, takes the average of the complexity scores of all majors linked to this occupation from the previous iteration, $k_{m,b-1}$. They are then used to compute the complexity index on the major side in the next iteration, $k_{m,b+1}$, which is the average of $k_{j,b}$ for all occupations linked to major m. The underlying idea is that occupations that require a complex set of skills are linked to, on average, majors that teach complex skills, while majors that equip more complex skills can send students to occupations that are more demanding. Thus, the complexity index has to be constructed recursively. Upon convergence, we obtain the

[4]The notion of network distance plays an important role here. In the context of our major-to-occupation network, majors that are closely connected share information more quickly than those further apart. In our example, the History major is only 2 hops away from the Education major as they are both connected to the Clerk occupation. However, for the History major to pass information to Engineering, the shortest path is 4 hops. However, given enough iterations, information from all connected nodes in the bipartite network is incorporated, thereby providing a robust and comprehensive measure of the complexity of skills associated with each major.

Major-Complexity-Index (MCI) on the major side, which is an accessibility-adjusted measure taking into account relevant neighboring majors that map students into the same occupations. The intuition is that by utilizing the major-to-occupation bipartite network, we are able to shed light on the underlying tripartite network connecting college majors to the skills they produce, and occupations to the skills they require, which is referred to as the "building-block" model according to Hidalgo and Hausmann (2009); Hausmann and Hidalgo (2011). In Appendix A, we elaborate on how a tripartite network of major-skill-occupation reduces down to a bipartite major-occupation network. The goal of the MCI is to infer the relative complexity of the skill set in each major based on the "building-block" model from the information contained within a bipartite flow network.

To implement this procedure, we need to a) specify the initial value of $k_{m,0}$, and b) construct an informative major-to-occupation network M. We provide a brief explanation of these preparatory steps below and leave the detailed discussion to Appendix B .

3.1 Initial Value

When applying the Method of Reflection to a major-to-occupation network, it is necessary to specify an initial value, $k_{m,0}$, for each major. We find that the specific initial values do not affect the standardized MCI following convergence. Provided that we start with an initial vector that ranks the majors according to skill specificity, the MCI invariably converges to the same value. Therefore, in our main analysis, we use the HHI of each major, constructed from the raw major-to-occupation network, as the initial value with no sensitivity of the results to this choice. In Appendix B.1, we show that the MCI initialized by the Gini coefficient and the share of Top-3 occupations consistently converge to the same values. This result closely relates to the interpretation of MCI as the eigenvector corresponding to the second eigenvalue of a major-to-major distance matrix, an idea introduced in Cristelli et al. (2013). We elaborate on this in detail in Appendix B.1.

3.2 Network by Revealed Comparative Advantage (RCA)

The most important ingredient of the recipe is an informative major-to-occupation network M, where $M_{mj} = 1$ if a link is recorded from a major m to an occupation j, and $M_{mj} = 0$ otherwise. To construct this binary matrix, following the original implementation by Hidalgo and Hausmann (2009), we compute the Revealed Comparative Advantage (RCA) of a major m for an occupation j as

$$RCA_{mj} = \frac{N_{mj}/N_m}{N_j/N},$$

where N_{mj} is the number of graduates from a major m who found a job in an occupation j, N_m is the total number of graduates from major m, N_j is the total number of graduates in occupation j, N is the total number of graduates in the network.

In the binary matrix M, a link between major m and occupation j is recorded (i.e., $M_{mj} = 1$) if $RCA_{mj} > 1$. Intuitively, a major m forms a link to an occupation j if and only if the fraction of its graduates placed in j is higher than the average across all majors. The matrix formulated by the RCA represents the relative value of a major by evaluating the comparative advantages it has over others in terms of job placements for its graduates. More explicitly, RCA for a major in a particular occupation is computed by comparing the share of graduates from this major in this occupation relative to the share of all graduates in this occupation. A high RCA indicates that a major has a comparative advantage in placing its graduates into a specific occupation, which could be interpreted as this major equipping its students with a specific set of skills highly demanded by this occupation. Therefore, the RCA matrix captures the unique relationships between majors and occupations, shedding light on the distinct skill sets learned from different majors and how they are valued in various occupations. This, in turn, enables us to create the Major Complexity Index (MCI) that reflects the complexity of skills taught in different majors based on the dynamics of a major-to-occupation network.

The RCA approach offers several advantages over alternatives. We provide an in-depth comparison with other approaches in Appendix B.2. Here we highlight two important features. First, one important caution in interpreting the Method of Reflections is the assumption that major-to-occupation mappings are primarily (if not entirely) based on the skills match. That is, for any major occupation pair, a link exists if and only if the required skills are matched. If some students, say by virtue of family-resource, end up in occupations that they do not qualify for in terms of skill requirements, it is then problematic to infer major values using job placements. The RCA matrix alleviates such a concern, and is notably robust against such a mismatch, by removing excess variation and focusing only on significant presences and absences (Hidalgo and Hausmann, 2009; Hidalgo, 2021; Balland et al., 2022). Second, the RCA is agnostic to the population size of nodes. If a fixed count were used as a trimming threshold, a larger major like Engineering would likely form links with every occupation, while a smaller major like Archaeology would potentially form no links at all. The RCA approach allows every major to have advantages in certain occupations and disadvantages in others, as it uses relative measures. This feature makes it a more robust and insightful measure for our purposes, as it concentrates on the most important information from a major-to-occupation network without being influenced by the sheer size of the majors.

3.3 Comparison to Other Major Indices

In the realm of higher education economics, there are several major-level indices designed to capture valuable insights about the skills taught across different academic disciplines. However, the MCI is distinctive in its approach, specifically leveraging the major-to-occupation network. Indices like the share of Top-3 occupations (Altonji et al., 2012) and Herfindahl-Hirschman Index (Blom et al., 2021) measure the concentration of occupational placements for each major. While they do measure the skill specificity accurately, both these measures ignore the nature of the occupations into which the majors place their students. That is, each occupation, in their analysis, is treated as a uniform, undifferentiated node. For example, in our data, both Computer Engineering and General Education exhibit a Herfindahl-Hirschman Index (HHI) of 0.35, which suggests that they are equally highly concentrated majors. However, a closer look at the specific occupations these majors feed into reveals a significant difference. Graduates from Computer Engineering are primarily concentrated in Computer Occupations, such as Programmers. In contrast, the majority of General

Education graduates become Preschool, Elementary, Middle, Secondary, and Special Education Teachers. Since Top-3 share and HHI do not take into account the links in the major-to-occupation network, they ignore such differences between the occupations into which different majors lead.

While it is intuitively evident that Computer Occupations and Teachers are markedly different, formulating a clear characterization of these differences can be challenging. The MCI solves this problem by characterizing occupations based on their linkages to the majors. For example, Computer Occupations are predominantly linked to majors such as Computer Science, Electrical Engineering, and Mathematics. Conversely, Teachers are linked to English Language and Literature or Liberal Arts. This approach leverages the inherent relationships between majors and occupations to gain insight into the nature of different career paths. Thus, the MCI provides a more nuanced representation by incorporating information about the occupations to which each major leads, an aspect often overlooked in other indices. Because majors and occupations mutually define each other in this paradigm, the calculation of the MCI naturally requires an iterative procedure. This iterative process allows for an increasingly useful extraction of information from the career outcomes associated with different majors.

The index based on the Gini coefficient, as put forth by Leighton and Speer (2020), represents a recent endeavor to define the specificity of majors by gauging the inequality in returns to education from different majors across different occupations. This index differs fundamentally from the MCI as it draws from a different data source (i.e. wage and salary) and concentrates particularly on the concept of major specificity. The calculation of the Gini coefficient requires income data for each occupation within every major, necessitating the use of an aggregate, coarse-grained categorization of majors and occupations. This is equivalent to presupposing that the major-to-occupation network is fully connected and there are no missing links between majors and occupations.[5] Meanwhile, the MCI capitalizes on the significant presences as well as absences, extracting valuable insights from when the links are missing. In subsequent sections, we demonstrate that the MCI exhibits stronger explanatory power than the Gini index in accounting for earning and employment

[5]For instance, if Nursing is in the major list, then every occupation needs to have some workers who graduated from Nursing.

differences across individuals. Nevertheless, when both indices are accounted for, neither loses its statistical significance, indicating that they capture distinct aspects of college majors that matter for the labor market outcomes.

Hemelt et al. (2021) describe majors by the skills they impart, exploiting online job advertisement data that contains both required skills and degree fields for occupations, and conclude that majors can be thought of as a portable bundle of skills. Although this approach of explicitly identifying skills embedded in each major holds numerous advantages, such data isn't readily available. On the other hand, the data requirements for calculating the MCI are relatively minimal. The MCI calculation necessitates only the observation of the graduates' occupational distribution for each major, without any additional information. This feature greatly enhances the practicability and accessibility of the MCI as a tool for characterizing and evaluating the value of different academic majors. We demonstrate and discuss the potential usage of MCI in subsequent sections.

4 Data

Our primary data is derived from the American Community Survey (ACS) covering the period from 2009-2019, which provides individual-level information on schooling choices (including college majors from 2009 onward) along with occupational outcomes, both of which are vital to creating a bipartite major-to-occupation network.[6] To learn about college majors, we carefully restrict our sample to individuals with only a Bachelor's degree, without double majors, who are currently not enrolled in school. Following Hemelt et al. (2021), we further limit our sample to full-time, full-year (FTFY) workers, who have worked at least 40 weeks a year and at least 30 hours a week in the past 12 months. This forms the basis for the MCI construction and earning analysis.

To construct the major-to-occupation network and subsequently the Major Complexity Index and other major specificity indexes, we follow Leighton and Speer (2020) and focus on individuals aged 25-35 to prioritize the occupational placements of recent graduates as well as to concentrate

[6]We use the ACS data obtained from IPUMS USA, University of Minnesota (`www.ipums.org`).

on skills obtained during collegiate studies, as opposed to post-graduate training.[7] Our MCI is primarily constructed from a network that consists of the detailed ACS major (*degfieldd*) and occupation (SOC4) categories.[8] To minimize noise resulting from spurious major-to-occupation linkages, we stipulate a minimum of 100 individuals per major and occupation. After applying these sample restrictions, we are left with 651,638 observations, enabling us to link 159 majors to 97 occupations.[9] The summary statistics of this sample are presented in Table 1, Panel A. Approximately half the sample is female, 80% are White, and around 9% are Hispanic. The potential experience, computed by subtracting 22 from age (as we only consider bachelor's degree earners), averages about 8 years, ranging from 3 to 13 years.

The earnings analysis is conducted on all individuals aged between 25 and 60 who have graduated in one of the 159 majors for which we have MCI and other major-specificity measures. We set the lower limit of the annual income (in 2010 dollars) at $500, leaving us with 1,999,140 observations, as demonstrated in Table 1, Panel B. This sample maintains a similar gender and race ratio to Panel A, though the proportion of Hispanic individuals is slightly lower (approximately 7.5%). The potential experience is higher on average by construction. The average annual income in this sample is about $71,199, ranging from $500 to $654,737 (with no top coding).

For the employment analysis, we eliminate the aforementioned bottom coding and Full-Time Full-Year (FTFY) restrictions and instead create an indicator variable denoting FTFY employment

[7]Using the MCI constructed from a major to occupation network with individuals aged 25-30 instead of 25-35 does not alter the results. See Appendix D Table D.1 for details.

[8]In ACS, majors are measured at both a coarse level (variable *degfield*), and a detailed level (variable *degfieldd*). The coarse level corresponds to the CIP2 level while the detailed level typically aligns with the CIP4 level. However, sometimes it falls between CIP2 and CIP4. For instance, the major number 2409 in degfieldd (Engineering Mechanics, Physics, and Science) can be mapped to three CIP4 majors - 14.11 (Engineering Mechanics), 14.12 (Engineering Physics), and 14.13 (Engineering Science) within the CIP2 category - 14 (Engineering). In some other cases, it is in-between CIP4 and CIP6. For instance, the major number 2311 in degfieldd (Social Science or History Teacher Education) can be mapped to two CIP6 majors - 13.1317 (Social Science Teacher Education) and 13.1328 (History Teacher Education) within the CIP4 category - 13.13 (Teacher Education and Professional Development, Specific Subject Areas). Using CIP4 in such a case would result in a loss of important heterogeneity, such as training of teachers across different subject areas. To minimize matching errors, we rely on ACS major variables, rather than matched CIP.

[9]At the CIP2 level, majors that are not in ACS include 28) reserve officer training corps (jrotc, rotc). 32) basic skills. 33) citizenship activities. 34) health-related knowledge and skills. 35) interpersonal and social skills. 36) leisure and recreational activities. 37) personal awareness and self-improvement. 53) high school/secondary diplomas and certificates. 60) residency programs.

status, applicable to individuals aged between 25 and 60, who graduated in one of the 159 majors we identified. Following these adjustments, our sample consists of 2,591,343 observations, as indicated in Table 1, Panel C. The sample characteristics align closely with those of Panel B, with around 80% of individuals working full-time.

To ensure the robustness of our main results and further comprehend the Major Complexity Indices, we utilize pooled data from the National Survey of Student Engagement (NSSE) spanning the years 2005 to 2011. This dataset provides rich student-level data, detailing the variety of assignments and tasks carried out across majors as well as information pertaining to student backgrounds. We calculate the major average of SAT scores, parental education, and the percentage of international students for each major, along with major characteristics surveyed from students such as the development of knowledge and skills through their college education and the number of hours spent on coursework. These computations are performed at the CIP2 level and are merged with our ACS sample. Overall, we manage to match 22 majors (out of 22 in the NSSE; out of 38 coarse-level majors in ACS) using the CIP2 between the two datasets. For additional details about the dataset and the matching process, please refer to Appendix F.

To bolster the robustness of our main results further, we integrate the MCI computed from the ACS with the National Longitudinal Survey of Youth 1997 (NLSY97) dataset, which collects extensive information about individuals' educational experiences and labor market outcomes together with important proxies of innate abilities and academic preparations prior to college. We successfully match 36 majors (out of 37 in the NLSY97; out of 38 coarse-level majors in ACS) using the CIP2 between the two datasets. More details about the dataset and the matching process can be found in Appendix G.

5 Results

The empirical results are organized into four subsections. We elaborate on the intuition of the MCI in Section 5.1, and present individual-level regression results for both earning and employment in Section 5.2. In section 5.3, we conduct a major-level analysis of MCI in predicting future income growth. Lastly, Section 5.4 investigates the relationship between the MCI and the detailed major characteristics from the NSSE.

5.1 MCI versus Previous Measures

As explained in the previous sections, based on the "building-block" model, the Major Complexity Index (MCI) constructs the complexity measure of acquired skills by recursively considering the complexity level of other majors that map into the same occupations. By doing so, the MCI incorporates information from the whole bipartite major-to-occupation flow network, and computes a comprehensive scalar measure of college majors that captures the latent skills taught in different majors and that are required by different occupations.

Figure 2 left panel displays the top 35 majors based on the occupational HHI (index before the iteration). Majors like nursing, education, engineering, computer science, and accounting have a high value of HHI which indicates these majors funnel a majority of their students into a relatively narrow range of occupations, suggesting a high level of specialization. They are regarded as highly specialized majors which are consistent with other major specificity measures like the Gini coefficient (Leighton and Speer, 2020) or Location Quotient (Hemelt et al., 2021). Meanwhile, the iteration method, founded on the major-to-occupation network, reclassifies the majors based on the similarity of occupation outcomes. The right panel in Figure 2 displays the top 15 and bottom 15 majors based on the MCI (index after the iteration), which updates the HHI by incorporating the information of other majors that map into the same occupations. Consequently, the ranking of similarly specific majors can drastically diverge.

Take two majors, Petroleum Engineering and Mathematics Teacher Education, for example. Despite the fact that these majors have remarkably similar HHI indices (0.44 and 0.45 respectively), the types of occupations their students find themselves in vary significantly. This distinction influences their major complexity calculation, resulting in considerable differences in ranking. In the case of the Petroleum Engineering major, students are mapped into occupations that are rarely accessible to other majors, such as Chemists and Materials Scientists, and consequently is connected to other specialized majors like Mechanical Engineering. Therefore, the Method of Reflections infers that the Petroleum Engineering major outputs students with skill sets that are hard to find elsewhere, and as such, it returns a high major complexity index rank (14/159). The reverse is true for the Mathematics Teacher Education major in which students are mapped into easily accessible occupations and thus connected to other relatively general majors, which in turn yields a lower MCI rank (125/159). The important takeaway here is that majors with a similar HHI can yield very different complexity rankings.

It is critical to understand that the MCI is not simply a refined measure of major specificity. Take, for instance, Accounting and Nursing – fields typically perceived as professionally oriented and hence often categorized as very specialized majors. In our MCI, Nursing ranks at the very bottom (157 out of 159), while Accounting sits in the 102nd position. As another illustration, Mathematics is commonly depicted as a general major in other indices (Leighton and Speer, 2020; Hemelt et al., 2021). Contrarily, in our MCI, it achieves a relatively high ranking of 37 out of 159. Please refer to Appendix C Table C.1-C.3 for the complete ranking of majors based on HHI (before iteration) and MCI (after iteration).

The MCI's unique character, separate from specificity measures, becomes apparent in Table 2, which shows the correlations between the MCI and other major specificity measures. As expected, due to their construction, HHI and Top3 are strongly correlated, and HHI and the Gini coefficient show less correlation, corroborating the findings of Leighton and Speer (2020). Despite being computed from the exact same data, the MCI (based on HHI as the initial values) and the HHI are nearly orthogonal, implying that the iterative processing of the RCA matrix captures crucial information about college majors that extend beyond the straightforward interpretation of specificity. Therefore, in deference to the original complexity literature, we refer to it as major complexity. Intuitively, a major's complexity can arise from multi-dimensional skills, the depth of each skill, and the interactions among those skills - aspects that the MCI implicitly captures without explicitly modeling the skill dimensionalities.

5.2 Earning and Employment Returns

Table 3 presents the OLS estimates of earnings and employment returns associated with the MCI and other indexes in Panel A and B, respectively. To facilitate interpretation and comparison, all major-level indexes are standardized to maintain a mean of zero and a variance of one across majors.[10] Columns (1)-(4) include one major index at a time while columns (5)-(7) add an additional index to the MCI, and column (8) controls for all four indexes (MCI, HHI, Top3 share, Gini coefficient). Across all specifications, gender, race, a quadratic function of potential experience, and year-fixed effects are controlled, with standard errors clustered at the major-year level.

Consider first the income returns in Panel A. Several features of the findings are noteworthy. First, a one standard deviation increase in the MCI raises salary by 8.19% in column (1) and the effect is statistically significant at the 1 percent level in explaining the earning differentials across individuals. In contrast, the income return to the HHI is only 2.57% per standard deviation. Second, the MCI notably increases the overall explanatory power of the model. The adjusted R^2 increases by 12% when we add the MCI on top of the HHI from column (2) to (5) (i.e., adjusted R^2 changes from 0.1224 to 0.1374); however, it only increases 3% when we add the HHI in addition to the MCI from column (1) to (5) (i.e., adjusted R^2 changes from 0.1335 to 0.1374). Third, the income returns to the Top3 share and Gini coefficient indexes are 4.31% and 7.62% per standard deviation, respectively, as shown in columns (3) and (4), and exhibit lower explanatory power (adjusted R^2) than the MCI in accounting for the earning differences. The returns decrease to 3.69% and 4.83% when the MCI is controlled for in columns (6) and (7). Fourth, the income return to the MCI remains robustly around 8-9% even when controlling for other major indexes as shown in columns (5)-(8), suggesting that the MCI captures distinct aspects of college majors that matter for wage and salary earnings. Lastly, consistent with the literature, we observe lower earning returns for females, non-whites, and Hispanics, and there is a diminishing return to potential experience.

[10] In the major complexity index, the Business Management and Administration major has a standardized index close to zero, while Physics is approximately one standard deviation above, and Marketing and Marketing Research is about one standard deviation below.

In Appendix D Table D.2, a robustness check is performed for the income regression controlling for occupation fixed effects at the SOC2 level. There are potentially two ways the skills acquired from a college major affect earning outcomes. One is that skills enable students to secure jobs in higher-paying occupations, and the other is that, within the same occupation, individuals with higher skills command higher wages. Interestingly, by controlling for occupation fixed effects, thereby eliminating the across-occupation channel, the return to the MCI decreases to approximately 4-5%, down from 8%, consistently across columns. This suggests that the wage return to the MCI can be broken down into a 3-4% component due to differences across occupations (at the SOC2 level), and a 4-5% component due to within-occupation variation. It's also worth noting that the returns to other major indexes, which use occupation information in the network, also diminish by about 23-33% using estimates from columns (2)-(7).

The discussion now turns to the employment returns in Table 3, Panel B.[11] We observe similar patterns for the major complexity index in explaining differences in employment across individuals. A one standard deviation increase in the MCI raises the probability of full-time work by 1.56% in column (1), which is more than double the return to the HHI (0.61%). These estimates remain statistically and economically significant in columns (5)-(8) when other major indexes are controlled for. The adjusted R^2 also increases when the MCI is added, although the increments are not as large as in the earnings regressions. Similarly, the probability of working full-time is lower for females, non-whites, and Hispanics, and diminishes over potential experience. Henceforth, we omit the estimates of these demographic controls in other tables.

We note another interesting observation: the explanatory power of the MCI increases with iterations when we apply the Method of Reflections. As shown in Appendix C Table C.4 Panel A, the income return increases from 0.67% to 8.19% between the 2nd and 50th iterations[12], and

[11] Our main results are based on the Linear Probability Model. The results are qualitatively similar when employing Logistic and Probit regressions, both of which exhibit the return to one standard deviation change in MCI to be roughly 2.5% increment in the FTFY employment.

[12] The estimated income return is 8.01% at the 20th iteration. We iterate further to the 50th iteration to ensure full convergence, meaning no further major ranking changes. See Appendix C Figure C.1 for the MCI ranking changes over iterations where majors with the largest changes are highlighted.

the adjusted R^2 increases by 11% from 0.1201 to 0.1335. This increase is surprisingly large considering the Method of Reflections is simply a manipulation on the same data. The estimated probability of working full-time in Panel B also increases from 0.23% to 1.56%, and the adjusted R^2 increases by 3.5% from 0.0398 to 0.0412. Intuitively, the complexity of a major arises from the number of marketable skills, the depth of each skill, and the interactions among those skills. The iterative process allows for an increasingly useful extraction of valuable information from the major-to-occupation flow matrix that is predictive of students' potential earning and employment outcomes without explicitly modeling skill dimensionality.

Aggregation

The estimate of returns to college majors can be sensitive to the major and occupation aggregation (Andrews et al., 2022). In our framework, different aggregations of majors and occupations create different networks from which we extract information. We investigate the robustness of our results to different aggregation schemes by computing four MCIs from different networks. These networks vary based on whether the major is coded at a coarse (variable *degfield* which corresponds to CIP2) or detailed level (variable *degfieldd* which mostly corresponds to CIP4) and whether the occupation is measured at a coarse (SOC2) or detailed level (SOC4).

Table 4 displays the earnings and employment estimates of MCIs derived from different networks.[13] First and foremost, it is evident that the explanatory power of MCIs for labor market outcome differentials is robust against varying levels of aggregation. Secondly, the variation in majors when they are defined at a more detailed level has a substantial impact on earnings and employment regression. For example, as shown in column (1), when using detailed major and detailed occupation, a one standard deviation increase in the MCI raises earnings by 8.19% and increases the probability of full-time employment by 1.56%. In contrast, these figures are lower, at 6.88% and 1.09% respectively, when using major at the coarse level with detailed occupation, as shown in column (3), with a lower adjusted R^2 as well. This pattern is similarly observed when comparing

[13] See Table D.3 for the correlation among the MCIs from different networks.

columns (2) to (4). Lastly, an interesting observation can be made by comparing columns (1) to (2). When detailed majors are paired with broader occupation groups, the coefficients and adjusted R^2 are larger in column (2) compared to when detailed majors are paired with detailed occupation categories in column (1). Meanwhile, when using coarse majors, the coefficient and adjusted R^2 are larger when paired with detailed occupation categories as shown in column (3) compared to coarse occupations in column (4). This suggests that a detailed level category from one side combined with a broad level from the other side yields the most information.

While the MCI computed from broad major categories[14] maintains a certain degree of consistency with rankings based on detailed majors (e.g., engineering and hard science majors tend to rank highly), crucial details and variances are missed in the former approach. For instance, when viewed through the lens of a coarse major category, the Education major ranks 34th out of 35. However, when we delve into the detailed major categorization, interesting nuances emerge. Science and Computer Teacher Education major ranks significantly higher (54th out of 159), contrasting starkly with Elementary Education, which ranks at the bottom (159th out of 159). These nuances underscore the heterogeneity within broader categories, shedding light on the varying complexity of skills within ostensibly similar fields of study. Such granularity cannot be achieved with some other major indices such as the Gini coefficient. The computation of the Gini coefficient requires a fully connected network, necessitating the use of coarse-grained, aggregated categories of majors and occupations.

Major Category

Table 5 delves into the robustness of the MCI's predictive power concerning labor market outcomes in comparison to traditional approaches. Traditional methodologies usually categorize majors by broad areas of study (like STEM vs. non-STEM, Arts and Humanities, Sciences, Social Sciences, etc.). Several interesting insights can be drawn from these comparisons.

First, as shown in Panel A column (2), earnings are on average significantly higher for STEM

[14] Results are available upon request.

majors. However, this effect becomes null when we control for the MCI, as demonstrated in column (3). Furthermore, the adjusted R^2 value increases by about 5% from 0.1270 to 0.1335, demonstrating the improved predictive power of our index. This suggests that the MCI is not simply a repackaging of traditional categorizations such as STEM vs non-STEM majors, although engineering and hard science majors tend to be highly ranked based on the MCI.

Second, as shown in Panel A, columns (5), even after accounting for the broad fields of study (8 categories with "Others" as the omitted group), majors with higher complexity scores still exhibit considerably higher average earnings.[15] Within a broad major field, detailed majors continue to vary in terms of their complexity scores. A one standard deviation increase in the MCI within a major group still results in an increase in income by 4.06%, which is roughly half of the estimate in column (1). Moreover, when we control for occupation fixed effects at the SOC2 level, the return further drops to 2.04% per standard deviation, but remains statistically significant at the 1% level, as shown in Table D.4. This implies that, even within the same major field and the same occupation group, majors with one standard deviation higher complexity still yield about 2% higher income returns.

Third, similar patterns can be observed for employment outcomes, as shown in Table 5 Panel B. However, it's noteworthy that the coefficient of STEM turns negative after controlling for the MCI in column (3).

Selection

In the analyses conducted thus far, there is an important caveat, particularly relevant in the context of education and labor economics, concerning selection bias. This arises when the type of students attracted to different majors inherently differs, thereby influencing labor market outcomes. For instance, if students from engineering and education majors exhibit differing labor market outcomes, this could be attributed to the distinct skill sets imparted by these two majors, but it could also be

[15] We note here, the adjusted R^2 in column (5) when controlling both MCI and broad major field is 0.1553 for income in Panel A, and 0.0446 for employment in Panel B. The corresponding figures are 0.1657 and 0.0458 when we control for major fixed effects at the detailed level instead, which is the dominant method used in the literature in identifying returns to college skill investments as proxied by college majors.

a function of the differing capabilities or preferences of the students these majors attract.

To address this concern, we undertake two robustness checks which are detailed in Appendix F and G. While establishing causality using observational data is always challenging, and therefore any claim of causality should be approached with caution, the results from these exercises provide some level of confidence that our findings are robust, and suggest that the MCI can capture meaningful variations across majors, beyond what can be accounted for by student selection effects.

The first exercise takes advantage of the information provided in the NSSE dataset. In this exercise, we merge major-level (at CIP2) characteristics such as average SAT verbal and math scores, the percentage of international students, as well as average education levels of both fathers and mothers from the NSSE into our ACS sample. The results controlling for these variables in this merged sample, displayed in Table F.1, are consistent with our main findings. In Panel A, the income return to the Major Complexity Index is 10.15% with basic controls in column (1). This figure drops to 8.25% when we control for major average SAT scores, parental education levels, and the percentage of international students in column (2). The employment return, as shown in Panel B, also maintains consistency across the two columns.

The second robustness check involves merging the MCI computed from the ACS with the NLSY97 dataset, which provides a more extensive set of observable controls at the individual level, particularly concerning preexisting abilities and high school preparations. It's important to note that, because this merge occurs at the CIP2 level from ACS to NLSY97, the MCI used is computed from the network between coarse majors (which corresponds to CIP2) and detailed occupation codes at SOC4. Although the NLSY97 sample is considerably smaller, the results, displayed in Table G.1, are nonetheless similar to our main results obtained from the ACS sample. With adjustment for various ability measures (SAT/ACT scores, AFQT score, and high school GPA) and household information (average household income during ages 15-19, parental education), we find that one standard deviation increase in the MCI is associated with roughly a 7-8% increase in salary and a 5% increase in the probability of working full-time. These estimates align closely with the one that only contains the basic controls in column (1). This exercise further reinforces our main findings, demonstrating that the MCI remains a significant predictor of labor market outcomes, even after adjusting for individual ability and household factors.

While the MCI demonstrates strong predictive power for labor market outcomes, it is critical to acknowledge that our findings rest on the selection on observables assumption, which is undoubtedly restrictive. Other factors unaccounted for in our model, such as individual interests, work-life balance preferences, risk tolerance, and other unmeasured characteristics can significantly influence students' major selection and subsequent labor market performance. If these unobserved variables were taken into account, the associations observed between the MCI and labor market outcomes could potentially be attenuated. Much of the recent efforts on returns to college majors have focused on identifying the causal impact of majors on future earnings, allowing for the selection on unobservables across majors (Kirkeboen et al., 2016; Hastings et al., 2013; Daly et al., 2022; Andrews et al., 2017; Bleemer and Mehta, 2022). To thoroughly isolate the complexity of the latent skills associated with each major, one would need either an extensive set of controls or additional assumptions, such as a complicated structural model (e.g., Arcidiacono 2004; Kinsler and Pavan 2015). Thus, while the MCI serves as a powerful predictor of labor market outcomes, the estimates derived from our analysis should be interpreted with due consideration. As such, they may be considered an upper bound on the true predictive power of the MCI.

Heterogeneity

Leighton and Speer (2020) reports that the returns to major specificity measures, including the Gini coefficient and other indexes, decrease with age. Similarly, the return to the major complexity may vary over the life cycle as well. To examine this, in Figure 3 and 4, we display the income and employment return to MCI over age, respectively. Specifically, they plot the coefficients obtained from the specification in Table 3 column (1) over different samples constructed with a sliding 10-year window of age (i.e., 25-35, 26-36, etc.). The income return at the early career is approximately 8.9% per standard deviation for the age group 25-35. This return gradually diminishes with age, and by the time individuals are between 50-60 years old, the return is around 7.5%. Similar patterns can be observed for employment return.

These observations suggest that the benefits of a complex major, as measured by the MCI, are most pronounced in the early- and mid-career. As workers age, the returns to major complexity tend to diminish, though they remain statistically significant. This could be due to a variety of factors such as career changes or skill obsolescence (Deming and Noray, 2020). However, despite this diminishing return, it is important to note that the MCI remains a significant predictor of both income and employment outcomes, even toward the end of one's career. This indicates that the skill bundle of college majors, as measured by the Major Complexity Index, can have long-lasting impacts on career outcomes, which aligns with the perspective that investment in higher education can yield advantages spanning potentially an entire professional career (Hoxby, 2020).

Table 6 delves deeper to examine potential heterogeneity in the effects of major complexity across different demographic groups, specifically by gender and race/ethnicity. Interestingly, the effects observed in Table 3 manifest differently across these subgroups. As column (2) indicates, males seem to reap higher returns to major complexity on average. All else being equal, a one standard deviation increase in the MCI boosts salary by 9.24% for males, a premium of 2.53% compared to females. This gender difference is statistically significant at the 1 percent level. This finding aligns with previous research that suggests that male graduates receive higher returns to education in some majors (e.g., Altonji et al. 2012; Hastings et al. 2013). When we look at racial/ethnic groups, there are observable differences as well. Whites appear to gain higher returns from major complexity, while the return is lower for Hispanics. Some studies have documented similar racial and ethnic disparities in labor market returns to college majors (e.g., Carnevale et al. 2013). These results suggest that the returns to major complexity may not be uniform across different demographic groups.

5.3 MCI Over time

The Economic Complexity Index (ECI), originally based on trade data, garnered considerable attention due to its strong correlation with countries' GDP growth (Hidalgo and Hausmann, 2009;

Hausmann et al., 2014; Hidalgo, 2021; Balland et al., 2022). Specifically, economies with higher ECIs tend to experience more rapid growth in the future, emphasizing the index's significance in capturing countries' production capacities and potentials. In our study, an analogous analysis would involve predicting future major-level mean income using the current major complexity level.

Our data from the ACS, however, is limited to the years 2009-2019, constraining our capacity to analyze long-term changes. Nevertheless, we can still utilize the available data for this exercise. We begin by constructing the MCI using data from each year.[16] Subsequently, using the major by year panel constructed, we perform a major-level analysis of the log mean income of major m, τ years into the future for $\tau = 1, 2, 5, 10$, with the current log mean income and the MCI as independent variables, controlling for female, White, and Hispanic ratios in the major, as:

$$logIncome_{m,t+\tau} = \beta_0 + \beta_1 \cdot MCI_{m,t} + \beta_2 \, logIncome_{m,t} + \beta_3 \, Ratio_{m,t} \; + u_{m,t+\tau}.$$

Our results in Table 7 demonstrate a correlation between the MCI and the future mean income associated with different majors. Interestingly, we notice that the coefficient of the MCI increases as we look into a longer time horizon (i.e., as τ increases) which suggests that the MCI is related to the long-term growth of college majors. For example, as shown in Panel A, the coefficient of the MCI on log income escalates from .0057 to .0284 when τ increases from 1 year in column (2) to 10 years in column (5). This indicates that the MCI not only correlates with the current returns on majors, as shown in column (1), but also predicts how these returns change over time.

In contrast, the HHI, while highly correlated with the current major average income, does not demonstrate a similar predictive power for future changes, as shown in Panel B. This differentiation further emphasizes the unique insights offered by the MCI in understanding both the current and future dynamics of income across college majors.

In Appendix E, we present changes in MCI rankings over time, using the first five years (2009-2013) and the last five years (2015-2019) of the data. The rankings remain mostly stable over this period, with a correlation of 0.93 between the MCIs at the major level. However, certain notable

[16] We restrict the majors and occupations so that each of them has at least 100 individuals each year. As a result, we have 156 majors instead of 159 in this exercise.

changes did occur. For example, the ranking for Information Science rose from 33 to 12. The ranking for Pharmacy, Pharmaceutical Sciences, and Administration also improved, rising from 112 to 62. On the other hand, the rankings for Commercial Art and Graphic Design declined from 73 to 114. Finance also dropped from 83 to 118, which might have been influenced by a decrease in job vacancies following the 2008 financial crisis. These observed changes in rankings over time may reflect underlying structural shifts in the labor market. Nonetheless, exploring these potential explanations in detail falls outside the scope of this paper. It would indeed be an intriguing avenue for future research to investigate these chronological patterns further.

5.4 MCI and Major Characteristics

Our preceding analyses suggest that the MCI reveals important aspects of college majors that matter to the earnings and employment of college graduates. In order to better understand what the MCI captures, we combine the MCI computed from the ACS with major-level characteristics from the National Survey of Student Engagement (NSSE) for the years 2005-2011. In total, we are able to match 22 majors (out of 22 majors in NSSE; out of 38 majors in ACS at *degfield*) at the CIP2 level between the two datasets. See Appendix F for the description of the NSSE dataset and the matching details.

Table 8 provides the pairwise correlation between the MCI (after iteration), HHI (before iteration), and a number of major-specific characteristics. First of all, in terms of students' academic preparation before coming to college, the MCI is positively correlated with students' performance in all three SAT measures (mathematics, verbal reasoning, and writing ability). Noticeably, the positive correlation between the MCI and SAT math score is particularly strong. In comparison, the correlation between the HHI and SAT variables is much weaker.

Interestingly, when examining areas where students report that their current academic programs have helped them develop further knowledge and skills, we see that further development of thinking critically and analytically, quantitative and complex problem solving, and the use of computing and information technology are positively correlated with the MCI measure. Similarly, analyzing the basic elements of an idea, experience, or theory, and applying theories or concepts to practical problems or in new situations also has a fairly strong positive correlation with the MCI. In contrast, there is little correlation between the MCI measure and the advancement of writing and speaking abilities in college. Taken together with the observed positive correlation with SAT verbal and writing scores, this suggests that higher MCI majors have students with high verbal and writing abilities (potentially developed prior to college), but who primarily report developing their ability to think critically and analytically, as well as the ability to analyze and solve quantitative, practical, and complex problems. Relatively speaking, the correlations between the HHI and those major features are much lower, except for the usage of computing and information technology.

Another important observation is the negative correlation between the MCI and broad general education ("Acquiring a broad general education", correlation -0.207). Meanwhile, the MCI does not demonstrate a correlation with the acquisition of job or work-related knowledge and skills ("Acquiring job or work-related knowledge and skills", 0.027). Meanwhile, high HHI majors tend to be more job-specific (0.132), and even less general (-0.287). This suggests that the MCI transcends the traditional distinction between general and job-specific skills. Presumably, the MCI's robustness in explaining earning and employment differentials is attributable more to the quantitative and analytical skills it encapsulates, rather than direct knowledge about job content.

We also find another intriguing set of correlations. Both the MCI and HHI are negatively associated with voting behavior, understanding self and others from diverse backgrounds, developing a personal code of values and ethics, spirituality, and contributions to the community. Different undergraduate degree programs adopt various approaches to prepare students for the labor market. Some are more vocationally oriented, such as engineering, while others adhere to the principles of a liberal arts education model (Ransom and Phipps, 2017). All these aforementioned elements represent significant values of higher education that go beyond job placements, and they are not particularly represented by the information contained in a major-to-occupation network. As Webber (2014) succinctly states, "While few would argue that a particular major should be chosen purely based on economic returns, the under-performing labor market and increasing tuition necessitate it to be at least considered by college underclassmen trying to decide their career path."

Regarding time commitment and effort, we observe that students in high MCI majors tend to report spending more hours preparing for classes, completing problem sets, and working on longer written assignments, more so than high HHI majors. If the time spent studying can be viewed as a proxy for coursework intensity and difficulty, this suggests that high MCI majors are more demanding on students and require them to invest greater efforts into their schooling. This increased investment, however, could potentially yield higher returns in the future.

6 Discussion

How are advanced skills formed through college education via different majors, and is this skill-production process responding to the changes in skill demand in the labor market (Conzelmann et al., 2023)? These are very hard questions to answer because skills are not directly observable, and some of them may not even be easily interpretable. In this paper, instead of explicitly modeling skill dimensionality (see, e.g., Kinsler and Pavan 2015; Hemelt et al. 2021), we take an alternative approach that computes a general measure of "complexity" for each major which reflects the skills taught in different majors and that are required by different occupations. This easily computable index adds another lens through which we can discuss these important questions.

Specifically, we apply the Method of Reflections, introduced by Hidalgo and Hausmann (2009) to parse through the major-to-occupation flow network and formulate a scalar measure that indicates the relative complexity of skills imparted by various majors. Our complexity measure appears to be a potent factor in explaining individual earning and employment differences across college majors, and the results remain robust to confounding factors and aggregation issues. Further results suggest that the MCI can not only account for current income disparities but also predict future major-level earning growth, a feat beyond the capability of typical specificity indices. Additional exercises reveal that the MCI strongly relates to advanced skills such as critical and analytical thinking, as well as abilities to analyze and solve quantitative, practical, and complex problems.

The major complexity index exhibit rankings of college majors that are naturally of interest to various stakeholders, including prospective students and their families in choosing college majors, as well as university administrators in charge of strategic planning of their schools. From students' perspective, it is essential to understand how the occupational outlook varies based on the choice of major. While it is well documented that expected earning is a key factor in choosing fields of study (Beffy et al., 2012; Wiswall and Zafar, 2015; Altonji et al., 2016), another equally important yet underexplored aspect is what occupations become available through the skills acquired in different college majors.

For administrators, the major complexity index presents a convenient and informative reference for their strategic planning. Evidently, colleges have been struggling to allocate resources across majors, particularly under budget-constrained circumstances. For instance, the University of Wisconsin at Stevens Point announced its elimination of 13 majors in 2018 to address "fiscal challenges" (Flaherty, 2018). Most recently, many universities are facing severe financial difficulties caused by the COVID-19 pandemic (e.g. Seltzer 2021), and forced to reallocate limited resources across majors. The major complexity index can facilitate this decision-making process by providing information on which majors prepare students with a more marketable combination of skills. It is worth emphasizing that the MCI is easily computable with a minimum data requirement. Many universities have a center for career development that conducts post-graduation surveys. This allows administrators to apply our proposed method to their own major-to-occupation network to produce individualized major complexity ranking. They can also compare it against the national or peer institutions' ranking of major complexity to better understand the comparative advantages of their own institution, and build the short- and long-term strategic plans accordingly.

One important caution in interpreting the Method of Reflections is the assumption that major-to-occupation mappings are primarily (if not entirely) based on the skills match. The RCA matrix adopted following Hidalgo and Hausmann (2009) alleviates such a concern by removing excess variation and focusing only on significant presences and absences. Despite this noted limitation, the Method of Reflections and the complexity index are widely connected to differences in economic growth, income inequality, gender inequality, humans development, and greenhouse emissions (see Hidalgo, 2021; Balland et al., 2022 for a detailed summary of its applications) due to the minimum data requirements and surprisingly strong explanatory power that this simple computation offers. Similarly, our generalized measure of MCI provides a useful tool for investigating difficult questions pertaining to the unobserved college skill production process. As aforementioned, a major-to-occupation network can be easily constructed from many data sources and one can utilize the rich information contained in a network structure to extract the major complexity feature, without relying on student-reported major characteristics. Moreover, we can exploit the dynamics of such a network over time to examine any structural changes within college education as a response to the changing nature of the labor market.

Several extensions of our analysis are in order. First, skill demand varies across local markets and time (Acemoglu and Autor, 2011; Deming and Kahn, 2018; Hershbein and Kahn, 2018; Deming and Noray, 2020; Hemelt et al., 2021). Our paper demonstrates the MCI dynamics over time by computing the MCI for each year. With appropriate data, one could also explore the spatial variations in major complexity. For instance, the MCI of a given major may vary across different types of institutions, such as public vs. private sector, varying college quality, as well as different geographic locations, etc. (Clotfelter 1999; Dale and Krueger 2002, 2014; Black and Smith 2004; Hoxby 2009; Ehrenberg 2012; Dillon and Smith 2020; among others). Second, in this paper, we focus on the complexity on the major side, while the complexity of required skills by occupation (i.e., Occupation Complexity Index - OCI) is actually computed as a byproduct of the Method of Reflections. One obvious and exciting extension is to analyze the OCI in various ways. For example, similar exercises to this paper can be performed to examine the return to occupation complexity over time, as well as the association between the occupation complexity measure and occupational information from rich data sources such as the O*NET. We leave analyses of these and related issues as possible directions for further research.

Reference

Acemoglu, D. and D. Autor (2011). Skills, tasks and technologies: Implications for employment and earnings. In *Handbook of labor economics*, Volume 4, pp. 1043–1171. Elsevier.

Altonji, J. G., P. Arcidiacono, and A. Maurel (2016). The analysis of field choice in college and graduate school: Determinants and wage effects. In *Handbook of the Economics of Education*, Volume 5, pp. 305–396. Elsevier.

Altonji, J. G., E. Blom, and C. Meghir (2012). Heterogeneity in human capital investments: High school curriculum, college major, and careers. *Annu. Rev. Econ. 4*(1), 185–223.

Andrews, R. J., S. A. Imberman, and M. F. Lovenheim (2017). Risky business? the effect of majoring in business on earnings and educational attainment. Technical report, National Bureau of Economic Research.

Andrews, R. J., S. A. Imberman, M. F. Lovenheim, and K. M. Stange (2022). The returns to college major choice: Average and distributional effects, career trajectories, and earnings variability. Technical report, National Bureau of Economic Research.

Arcidiacono, P. (2004). Ability sorting and the returns to college major. *Journal of econometrics 121*(1-2), 343–375.

Balland, P.-A., T. Broekel, D. Diodato, E. Giuliani, R. Hausmann, N. O'Clery, and D. Rigby (2022). Reprint of the new paradigm of economic complexity. *Research Policy 51*(8), 104568.

Beffy, M., D. Fougere, and A. Maurel (2012). Choosing the field of study in postsecondary education: Do expected earnings matter? *Review of Economics and Statistics 94*(1), 334–347.

Black, D. A. and J. A. Smith (2004). How robust is the evidence on the effects of college quality? evidence from matching. *Journal of econometrics 121*(1-2), 99–124.

Bleemer, Z. and A. Mehta (2022). Will studying economics make you rich? a regression discontinuity analysis of the returns to college major. *American Economic Journal: Applied Economics 14*(2), 1–22.

Blom, E., B. C. Cadena, and B. J. Keys (2021). Investment over the business cycle: Insights from college major choice. *Journal of Labor Economics 39*(4), 1043–1082.

Carnevale, A. P., J. Strohl, and M. Melton (2013). What's it worth?: The economic value of college majors.

Clotfelter, C. T. (1999). The familiar but curious economics of higher education: Introduction to a symposium. *Journal of Economic Perspectives 13*(1), 3–12.

Conzelmann, J., S. W. Hemelt, B. Hershbein, S. Martin, A. Simon, and K. Stange (2023). Skills, majors, and jobs: Does higher education respond? In *NBER economics of education program meeting spring 2023*. APPAM.

Cristelli, M., A. Gabrielli, A. Tacchella, G. Caldarelli, and L. Pietronero (2013). Measuring the intangibles: A metrics for the economic complexity of countries and products. *PloS one 8*(8), e70726.

Cunha, F. and J. Heckman (2007). The technology of skill formation. *American Economic Review 97*(2), 31–47.

Cunha, F., J. J. Heckman, L. Lochner, and D. V. Masterov (2006). Interpreting the evidence on life cycle skill formation. *Handbook of the Economics of Education 1*, 697–812.

Dale, S. B. and A. B. Krueger (2002). Estimating the payoff to attending a more selective college: An application of selection on observables and unobservables. *The Quarterly Journal of Economics 117*(4), 1491–1527.

Dale, S. B. and A. B. Krueger (2014). Estimating the effects of college characteristics over the career using administrative earnings data. *Journal of human resources 49*(2), 323–358.

Daly, M., M. F. Jensen, and D. le Maire (2022). University admission and the similarity of fields of study: Effects on earnings and skill usage. *Labour Economics 75*, 102118.

Deming, D. and L. B. Kahn (2018). Skill requirements across firms and labor markets: Evidence from job postings for professionals. *Journal of Labor Economics 36*(S1), S337–S369.

Deming, D. J. (2017). The growing importance of social skills in the labor market. *The Quarterly Journal of Economics 132*(4), 1593–1640.

Deming, D. J. and K. Noray (2020). Earnings dynamics, changing job skills, and stem careers. *The Quarterly Journal of Economics 135*(4), 1965–2005.

Dillon, E. W. and J. A. Smith (2020). The consequences of academic match between students and colleges. *Journal of Human Resources 55*(3), 767–808.

Ehrenberg, R. G. (2012). American higher education in transition. *Journal of Economic Perspectives 26*(1), 193–216.

Flaherty, C. (2018). U wisconsin-stevens point to eliminate 13 majors. *Inside Higher Ed*.

Hamermesh, D. S. and S. G. Donald (2008). The effect of college curriculum on earnings: An affinity identifier for non-ignorable non-response bias. *Journal of Econometrics 144*(2), 479–491.

Hastings, J. S., C. A. Neilson, and S. D. Zimmerman (2013). Are some degrees worth more than others? evidence from college admission cutoffs in chile. Technical report, National Bureau of Economic Research.

Hausmann, R. and C. A. Hidalgo (2011). The network structure of economic output. *Journal of Economic Growth 16*(4), 309–342.

Hausmann, R., C. A. Hidalgo, S. Bustos, M. Coscia, and A. Simoes (2014). *The atlas of economic complexity: Mapping paths to prosperity*. Mit Press.

Hemelt, S. W., B. Hershbein, S. M. Martin, and K. M. Stange (2021). College majors and skills: Evidence from the universe of online job ads. Technical report, National Bureau of Economic Research.

Hershbein, B. and L. B. Kahn (2018). Do recessions accelerate routine-biased technological change? evidence from vacancy postings. *American Economic Review 108*(7), 1737–1772.

Hidalgo, C. A. (2021). Economic complexity theory and applications. *Nature Reviews Physics*, 1–22.

Hidalgo, C. A. and R. Hausmann (2009). The building blocks of economic complexity. *Proceedings of the national academy of sciences 106*(26), 10570–10575.

Hoxby, C. M. (2009). The changing selectivity of american colleges. *Journal of Economic perspectives 23*(4), 95–118.

Hoxby, C. M. (2020). The productivity of us postsecondary institutions. In *Productivity in Higher Education*, pp. 31–66. University of Chicago Press.

Kinsler, J. and R. Pavan (2015). The specificity of general human capital: Evidence from college major choice. *Journal of Labor Economics 33*(4), 933–972.

Kirkeboen, L. J., E. Leuven, and M. Mogstad (2016). Field of study, earnings, and self-selection. *The Quarterly Journal of Economics 131*(3), 1057–1111.

Kuhn, P. and C. Weinberger (2005). Leadership skills and wages. *Journal of Labor Economics 23*(3), 395–436.

Leighton, M. and J. D. Speer (2020). Labor market returns to college major specificity. *European Economic Review 128*, 103489.

Light, A. and S. Schreiner (2019). College major, college coursework, and post-college wages. *Economics of Education Review 73*, 101935.

Lovenheim, M. F. and J. Smith (2023). Returns to different postsecondary investments: Institution type, academic programs, and credentials. In *Handbook of the Economics of Education (Forthcoming)*.

Martin, S. (2022). College major specificity, earnings growth, and job changing. Technical report, Working Paper.

Ransom, M. R. and A. Phipps (2017). The changing occupational distribution by college major. In *Skill Mismatch in Labor Markets*, pp. 129–171. Emerald Publishing Limited.

Seltzer, R. (2021). N.J. university could cut 26% of full-time faculty amid budget woes. *Inside Higher Ed*.

Silos, P. and E. Smith (2015). Human capital portfolios. *Review of Economic Dynamics 18*(3), 635–652.

Webber, D. A. (2014). The lifetime earnings premia of different majors: Correcting for selection based on cognitive, noncognitive, and unobserved factors. *Labour economics 28*, 14–23.

Wiswall, M. and B. Zafar (2015). Determinants of college major choice: Identification using an information experiment. *The Review of Economic Studies 82*(2), 791–824.

Figures and Tables

Ranking	Major Name	HHI	Ranking	Major Name	MCI
1	nursing	0.71	1	computer science	2.09
2	teacher education: multiple levels	0.54	2	electrical engineering	2.07
3	special needs education	0.51	3	computer engineering	1.86
4	elementary education	0.49	4	materials science	1.84
5	**mathematics teacher education**	0.45	5	chemical engineering	1.79
6	**petroleum engineering**	0.44	6	civil engineering	1.73
7	science and computer teacher education	0.41	7	architectural engineering	1.66
8	early childhood education	0.41	8	mechanical engineering	1.63
9	computer science	0.39	9	naval architecture and marine engineering	1.61
10	art and music education	0.39	10	engineering mechanics, physics, and science	1.60
11	nuclear, industrial radiology, and biological technologies	0.38	11	mining and mineral engineering	1.58
12	secondary teacher education	0.37	12	construction services	1.52
13	language and drama education	0.36	13	miscellaneous engineering technologies	1.51
14	medical technologies technicians	0.35	14	**petroleum engineering**	1.49
15	general education	0.35	15	general engineering	1.48
16	computer engineering	0.35	.	.	.
17	medical assisting services	0.34	.	.	.
18	accounting	0.32	125	**mathematics teacher education**	-0.88
19	civil engineering	0.30	.	.	.
20	computer programming and data processing	0.29	.	.	.
21	mechanical engineering	0.29	145	art history and criticism	-1.33
22	computer and information systems	0.28	146	psychology	-1.33
23	social science or history teacher education	0.25	147	communications	-1.37
24	aerospace engineering	0.25	148	teacher education: multiple levels	-1.39
25	construction services	0.24	149	special needs education	-1.40
26	actuarial science	0.24	150	communication disorders sciences and services	-1.45
27	information sciences	0.24	151	english language and literature	-1.53
28	social work	0.23	152	visual and performing arts	-1.55
29	chemical engineering	0.23	153	journalism	-1.56
30	computer information management and security	0.22	154	social work	-1.62
31	mining and mineral engineering	0.22	155	family and consumer sciences	-1.65
32	educational psychology	0.22	156	early childhood education	-1.70
33	materials engineering and materials science	0.21	157	nursing	-1.76
34	materials science	0.21	158	language and drama education	-1.83
35	pharmacy, pharmaceutical sciences, and administration	0.21	159	elementary education	-2.21

Figure 2: Ranking Changes over Iterations. The left panel in the figure displays the top 35 majors based on the occupational HHI (index before the iteration). The right panel in the figure presents the top 15 and bottom 15 majors based on the MCI (index after the iteration), which updates the HHI by incorporating the information of other majors that map into the same occupations.

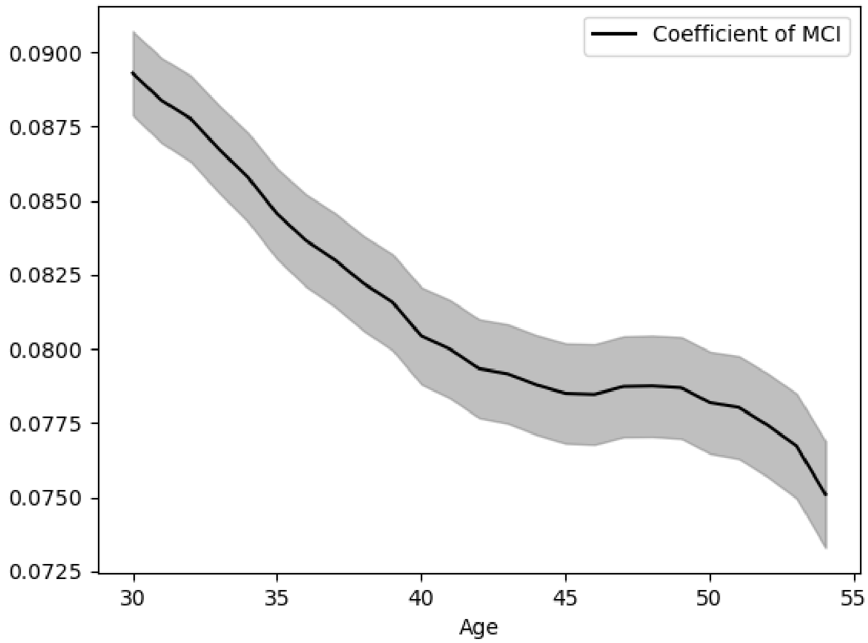

Figure 3: Income Return over Age Window. The figure plot point estimate of the coefficient on MCI obtained from the basic specification in Table 3, Panel A, Column (1) over different samples constructed with a sliding 10-year window of age (i.e., 25-35, 26-36, etc.). The gray shade represents the 95% confidence interval.

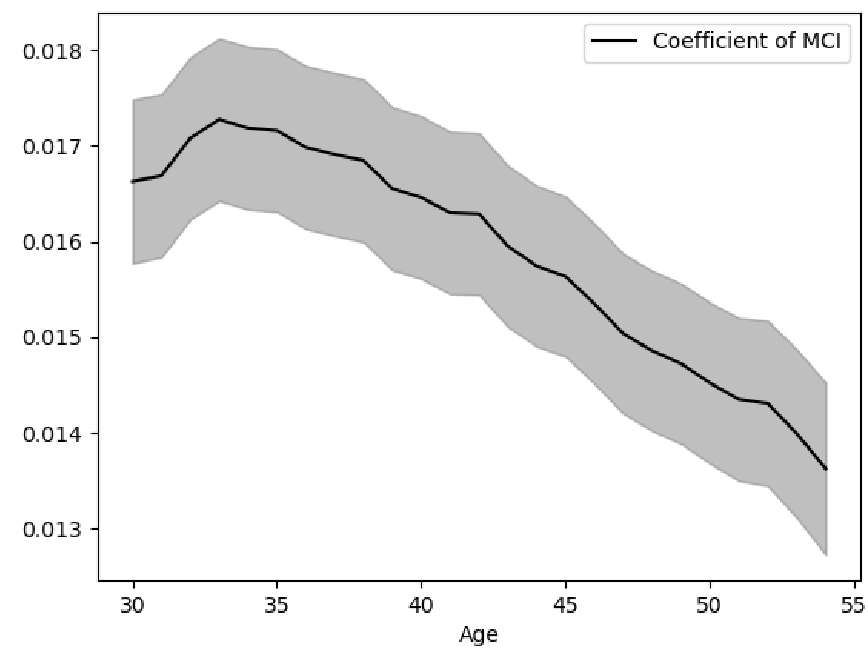

Figure 4: Employment Return over Age Window. The figure plot the point estimate with a 95% confidence interval obtained from the basic specification in Table 3, Panel B, Column (1) over different samples constructed with a sliding 10-year window of age (i.e., 25-35, 26-36, etc.).

Table 1: Summary Statistics

	Mean	Std. Dev.	Min	Max	N
Panel A: MCI Construction					
Female	0.490	0.499	0	1	651,638
White	0.800	0.399	0	1	651,638
Hispanic	0.090	0.286	0	1	651,638
Experience	7.983	3.130	3	13	651,638
Panel B: Earning Analysis					
Female	0.472	0.499	0	1	1,999,140
White	0.811	0.391	0	1	1,999,140
Hispanic	0.075	0.263	0	1	1,999,140
Experience	20.084	10.267	3	38	1,999,140
Annual Income (2010 $)	71,199	63,507	500	654,737	1,999,140
Panel C: Employment Analysis					
Female	0.514	0.499	0	1	2,591,343
White	0.811	0.391	0	1	2,591,343
Hispanic	0.075	0.264	0	1	2,591,343
Experience	20.458	10.360	3	38	2,591,343
Full-time Working Status	0.801	0.398	0	1	2,591,343

Note: Panel A presents the summary statistics for the dataset utilized in constructing both the major-to-occupation network and subsequent Major Complexity Index, along with other indexes related to major specificity; Panel B provides the summary statistics for the dataset employed in the earnings analysis, where the annual earning is adjusted for inflation to 2010 dollars; and Panel C reports the summary statistics for the sample used in the employment analysis where full-time working status is 1 if one has worked at least 40 weeks a year and at least 30 hours a week in the past 12 months. Source: American Community Survey (ACS), 2009-2019.

Table 2: Correlation Among Indices

	HHI	TOP3	GINI	MCI
HHI	1.0000			
TOP3	0.9240	1.0000		
GINI	0.1406	0.1261	1.0000	
MCI	0.0117	0.2175	0.0657	1.0000

Note: All major-level indexes are computed from a network connecting detailed ACS majors (which mostly correspond to CIP4) and detailed occupation categories (SOC4), and are standardized to have a mean of zero and a variance of one across 159 detailed ACS majors.

Table 3: Earning and Employment Return on the Major Complexity Index (MCI)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: DV - log of Annual Income (2010 $)								
MCI	0.0819***				0.0864***	0.0777***	0.0809***	0.0836***
	(0.0060)				(0.0045)	(0.0056)	(0.0061)	(0.0060)
HHI		0.0257***			0.0320***			0.0218
		(0.0050)			(0.0053)			(0.0165)
TOP3			0.0431***			0.0369***		0.0124
			(0.0056)			(0.0062)		(0.0169)
GINI				0.0762***			0.0483**	0.0003
				(0.0180)			(0.0207)	(0.0163)
Female	-0.2488***	-0.3296***	-0.3254***	-0.3162***	-0.2602***	-0.2591***	-0.2489***	-0.2601***
	(0.0053)	(0.0060)	(0.0064)	(0.0077)	(0.0038)	(0.0039)	(0.0052)	(0.0037)
White	0.1128***	0.1011***	0.1043***	0.0983***	0.1182***	0.1178***	0.1131***	0.1181***
	(0.0069)	(0.0063)	(0.0063)	(0.0065)	(0.0057)	(0.0059)	(0.0067)	(0.0057)
Hispanic	-0.1658***	-0.1637***	-0.1643***	-0.1645***	-0.1642***	-0.1651***	-0.1654***	-0.1645***
	(0.0045)	(0.0046)	(0.0046)	(0.0045)	(0.0045)	(0.0045)	(0.0045)	(0.0046)
Experience	0.0513***	0.0522***	0.0520***	0.0523***	0.0512***	0.0512***	0.0513***	0.0512***
	(0.0007)	(0.0007)	(0.0007)	(0.0007)	(0.0007)	(0.0007)	(0.0007)	(0.0007)
Experience^2	-0.0009***	-0.0010***	-0.0010***	-0.0010***	-0.0010***	-0.0010***	-0.0009***	-0.0010***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared Adj.	0.1335	0.1224	0.1252	0.1206	0.1374	0.1373	0.1338	0.1375
Panel B: DV - Full-time Working Status								
MCI	0.0156***				0.0170***	0.0150***	0.0157***	0.0187***
	(0.0012)				(0.0010)	(0.0011)	(0.0013)	(0.0015)
HHI		0.0061***			0.0077***			0.0131***
		(0.0010)			(0.0010)			(0.0032)
TOP3			0.0090***			0.0082***		-0.0056
			(0.0012)			(0.0013)		(0.0037)
GINI				0.0026			-0.0027	-0.0158***
				(0.0044)			(0.0046)	(0.0034)
Female	-0.1358***	-0.1516***	-0.1504***	-0.1487***	-0.1382***	-0.1379***	-0.1358***	-0.1385***
	(0.0014)	(0.0017)	(0.0017)	(0.0018)	(0.0014)	(0.0014)	(0.0014)	(0.0014)
White	0.0078***	0.0054***	0.0059***	0.0048***	0.0089***	0.0087***	0.0078***	0.0089***
	(0.0021)	(0.0018)	(0.0018)	(0.0019)	(0.0019)	(0.0019)	(0.0021)	(0.0019)
Hispanic	-0.0111***	-0.0105***	-0.0107***	-0.0108***	-0.0108***	-0.0110***	-0.0111***	-0.0107***
	(0.0013)	(0.0013)	(0.0013)	(0.0013)	(0.0013)	(0.0013)	(0.0013)	(0.0014)
Experience	0.0046***	0.0047***	0.0047***	0.0047***	0.0045***	0.0045***	0.0046***	0.0045***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Experience^2	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared Adj.	0.0412	0.0402	0.0405	0.0398	0.0420	0.0418	0.0412	0.0421

Note: The dependent variable in Panel A is the natural log of annual income in 2010 dollars, with 1,999,140 observations. The dependent variable in Panel B is the full-time working status with 2,591,343 observations. Gender, White, Hispanic, quadratic function of potential experience, and year-fixed effects are controlled for in all specifications. The MCI and other indexes are computed from a network connecting detailed ACS majors (which mostly correspond to CIP4) and detailed occupation categories (SOC4). Standard errors are clustered at the major-year level. Significance is as follows: one-percent=***, five-percent=**, and ten-percent=*.

Table 4: Earning and Employment Return on the MCI - Aggregation

Network Structure	(1) Major (Detailed) Occupation (Detailed)	(2) Major (Detailed) Occupation (Coarse)	(3) Major (Coarse) Occupation (Detailed)	(4) Major (Coarse) Occupation (Coarse)
Panel A: DV - log of Annual Income (2010$)				
$\text{MCI}_{SOC4}^{DetailedMajor}$	0.0819***			
	(0.0060)			
$\text{MCI}_{SOC2}^{DetailedMajor}$		0.1009***		
		(0.0060)		
$\text{MCI}_{SOC4}^{CoarseMajor}$			0.0688***	
			(0.0047)	
$\text{MCI}_{SOC2}^{CoarseMajor}$				0.0514***
				(0.0050)
R-squared Adj.	0.1335	0.1374	0.1308	0.1259
Panel B: DV - Status of Working Full-Time				
$\text{MCI}_{SOC4}^{DetailedMajor}$	0.0156***			
	(0.0012)			
$\text{MCI}_{SOC2}^{DetailedMajor}$		0.0174***		
		(0.0012)		
$\text{MCI}_{SOC4}^{CoarseMajor}$			0.0109***	
			(0.0009)	
$\text{MCI}_{SOC2}^{CoarseMajor}$				0.0054***
				(0.0010)
R-squared Adj.	0.0412	0.0413	0.0406	0.0400

Note: The dependent variable in Panel A is the natural log of annual income in 2010 dollars, with 1,999,140 observations. The dependent variable in Panel B is the full-time working status with 2,591,343 observations. Gender, White, Hispanic, quadratic function of potential experience, and year-fixed effects are controlled for in all specifications. The MCIs across columns are computed from different networks. These networks vary based on whether the major is coded at a coarse (ACS variable *degfield* which corresponds to CIP2) or detailed level (ACS variable *degfieldd* which mostly corresponds to CIP4) and whether the occupation is measured at a coarse (SOC2) or detailed level (SOC4). Standard errors are clustered at the major-year level. Significance is as follows: one-percent=***, five-percent=**, and ten-percent=*.

Table 5: Earning and Employment Return on the MCI - Controlling for Major Categories

	(1)	(2)	(3)	(4)	(5)
Panel A: DV - log of Annual Income (2010 $)					
MCI	0.0819***		0.0805***		0.0406***
	(0.0060)		(0.0093)		(0.0071)
STEM		0.1455***	0.0048		
		(0.0124)	(0.0166)		
Arts and Humanities				0.0259*	0.0554***
				(0.0135)	(0.0134)
Business				0.2109***	0.2148***
				(0.0135)	(0.0125)
Education				-0.1426***	-0.0982***
				(0.0113)	(0.0115)
Engineering				0.3696***	0.3006***
				(0.0130)	(0.0168)
Professional				0.2357***	0.2726***
				(0.0154)	(0.0175)
Science				0.1763***	0.1354***
				(0.0187)	(0.0134)
Social Science				0.1203***	0.1444***
				(0.0210)	(0.0187)
R-squared Adj.	0.1335	0.1270	0.1335	0.1541	0.1553
Panel B: DV - Full-time Working Status					
MCI	0.0156***		0.0179***		0.0074***
	(0.0012)		(0.0020)		(0.0012)
STEM		0.0231***	-0.0081**		
		(0.0021)	(0.0037)		
Arts and Humanities				-0.0388***	-0.0340***
				(0.0046)	(0.0042)
Business				0.0319***	0.0322***
				(0.0041)	(0.0036)
Education				-0.0249***	-0.0171***
				(0.0047)	(0.0043)
Engineering				0.0348***	0.0218***
				(0.0041)	(0.0042)
Professional				0.0212***	0.0275***
				(0.0053)	(0.0053)
Science				0.0159***	0.0082**
				(0.0043)	(0.0039)
Social Science				-0.0166***	-0.0126***
				(0.0046)	(0.0040)
R-squared Adj.	0.0412	0.0403	0.0413	0.0445	0.0446

Note: The dependent variable in Panel A is the natural log of annual income in 2010 dollars, with 1,999,140 observations. The dependent variable in Panel B is the full-time working status with 2,591,343 observations. Gender, White, Hispanic, quadratic function of potential experience, and year-fixed effects are controlled for in all specifications. The MCI is computed from a network connecting detailed ACS majors (which mostly correspond to CIP4) and detailed occupation categories (SOC4). Standard errors are clustered at the major-year level. Significance is as follows: one-percent=***, five-percent=**, and ten-percent=*.

Table 6: Earning and Employment Return on the MCI - Heterogeneity

	(1)	(2)	(3)	(4)	(5)
Panel A: DV - log of Annual Income (2010$)					
MCI	0.0819***	0.0924***	0.0659***	0.0848***	0.0801***
	(0.0060)	(0.0042)	(0.0094)	(0.0062)	(0.0068)
MCI *Female		-0.0253**			-0.0248**
		(0.0100)			(0.0099)
MCI *White			0.0205***		0.0190***
			(0.0061)		(0.0060)
MCI *Hispanic				-0.0396***	-0.0382***
				(0.0033)	(0.0033)
R-squared Adj.	0.1335	0.1338	0.1337	0.1337	0.1342
Panel B: DV - Status of Working Full-Time					
MCI	0.0156***	0.0162***	0.0006	0.0163***	0.0019
	(0.0012)	(0.0009)	(0.0021)	(0.0012)	(0.0016)
MCI*Female		-0.0014			-0.0012
		(0.0022)			(0.0022)
MCI*White			0.0192***		0.0190***
			(0.0016)		(0.0016)
MCI*Hispanic				-0.0093***	-0.0082***
				(0.0011)	(0.0011)
R-squared Adj.	0.0412	0.0412	0.0417	0.0413	0.0417

Note: The dependent variable in Panel A is the natural log of annual income in 2010 dollars, with 1,999,140 observations. The dependent variable in Panel B is the full-time working status with 2,591,343 observations. Gender, White, Hispanic, quadratic function of potential experience, and year-fixed effects are controlled for in all specifications. The MCI is computed from a network connecting detailed ACS majors (which mostly correspond to CIP4) and detailed occupation categories (SOC4). Standard errors are clustered at the major-year level. Significance is as follows: one-percent=***, five-percent=**, and ten-percent=*.

Table 7: Major-level Analysis of the MCI in Predicting Future Income

	(1) $\log(\text{Income})_{m,t}$	(2) $\log(\text{Income})_{m,t+1}$	(3) $\log(\text{Income})_{m,t+2}$	(4) $\log(\text{Income})_{m,t+5}$	(5) $\log(\text{Income})_{m,t+10}$
Panel A: MCI					
$\text{MCI}_{m,t}$	0.0624***	0.0057*	0.0087**	0.0208***	0.0284**
	(0.0063)	(0.0032)	(0.0036)	(0.0052)	(0.0135)
$\log(\text{Income})_{m,t}$		0.8739***	0.8519***	0.8208***	0.7554***
		(0.0123)	(0.0139)	(0.0181)	(0.0468)
R-squared Adj.	0.5243	0.8880	0.8720	0.8644	0.8598
No. observations	1,716	1,560	1,404	936	156
No. years	11	10	9	6	1
Panel B: HHI					
$\text{HHI}_{m,t}$	0.0215***	0.0016	-0.0009	-0.0027	-0.0111
	(0.0043)	(0.0022)	(0.0025)	(0.0032)	(0.0078)
$\log(\text{Income})_{m,t}$		0.8776***	0.8601***	0.8444***	0.8049***
		(0.0121)	(0.0137)	(0.0177)	(0.0447)
R-squared Adj.	0.5041	0.8878	0.8715	0.8621	0.8576
No. observations	1,716	1,560	1,404	936	156
No. years	11	10	9	6	1

Note: The MCI and HHI are constructed using yearly data from ACS, 2009-201, from a network connecting detailed ACS majors (which mostly correspond to CIP4) and detailed occupation categories (SOC4). In this exercise, we have only 156 (out of 159 majors) that have at least 100 individuals in each major and occupation each year. Column (1) reports the major-level analysis of the log mean income for major m in year t on the current MCI (in Panel A) and HHI (in Panel B), controlling for the female, White, and Hispanic ratios in major m year t. Column (2)-(5) presents the major-level analysis of the log mean income for major m, τ years into the future for $\tau = 1,2,5,10$, shown in each column respectively, on the current MCI (in Panel A) and HHI (in Panel B), with the current log mean income of major m in year t as well as the female, White, and Hispanic ratios of major m in year t controlled for. The number of observations changes across columns depending on how many years are utilized in the analysis. Standard errors are shown in parentheses. Significance is as follows: one-percent=***, five-percent=**, and ten-percent=*.

Table 8: Pairwise Correlations Between the MCI and NSSE Major Specific Characteristics

Variable Description	MCI	HHI
Standardized Test Scores		
SAT Verbal	0.105	0.038
SAT Mathematics	0.226	0.112
SAT Writing	0.091	-0.024
SAT Total	0.190	0.087
SAT Total (ACT converted)	0.188	0.085
Student Report - Developed Knowledge and Skills		
Writing clearly and effectively	0.047	-0.016
Speaking clearly and effectively	0.007	-0.053
Acquiring a broad general education	-0.207	-0.287
Acquiring job or work-related knowledge and skills	0.027	0.132
Thinking critically and analytically	0.274	0.092
Analyzing quantitative problems	0.260	0.140
Solving complex real-world problems	0.392	0.229
Using computing and information technology	0.193	0.244
Working effectively with others	0.160	0.158
Learning effectively on your own	0.161	-0.034
Memorizing facts, ideas, or methods from your courses and readings	-0.269	-0.160
Analyzing the basic elements of an idea, experience, or theory	0.311	-0.023
Synthesizing and organizing ideas, information, or experiences	0.148	-0.004
Making judgments about the value of information, arguments, or methods	0.085	0.029
Applying theories or concepts to practical problems or in new situations	0.266	0.148
Voting in local, state, or national elections	-0.095	-0.163
Understanding yourself	-0.232	-0.232
Understanding people of other racial and ethnic backgrounds	-0.186	-0.229
Developing a personal code of values and ethics	-0.095	-0.161
Developing a deepened sense of spirituality	-0.329	-0.186
Contributing to the welfare of your community	-0.131	-0.175
Student Report - Time Spent		
Hours Spent Preparing for class	0.100	0.020
Amount of problem sets that take more than an hour to complete	0.219	0.175
Amount of problem sets that take less than an hour to complete	-0.024	-0.140
Number of written papers or reports: 20 pages or more	0.564	0.366
Number of written papers or reports: between 5 and 19 pages	0.181	-0.001
Number of written papers or reports: fewer than 5 pages	0.136	-0.131
Challenging exams during the current school year	0.194	0.099

Note: Correlation based on variations across 22 matched majors (out of 22 majors in NSSE; out of 38 majors in ACS) at the CIP2 level between the two datasets. The MCI and HHI are thus computed from a network connecting coarse-level ACS majors (which correspond to CIP2) and detailed occupation categories (SOC4).

Appendix A Building Block Model and Flow Network

In this section, we briefly explain the intuition of the "building block model" (introduced in Hidalgo and Hausmann 2009 and solved analytically in Hausmann and Hidalgo 2011) in the context of a major-to-occupation flow network. For a more detailed discussion, see Hidalgo (2021).

Using the same example in Section 3 (Figure 1), there are four majors and four occupations. Now suppose the matching of students between majors and occupations is based on five latent skills. On the left-hand-side of Figure A.1, a link between a major and a skill indicates that this major has a comparative advantage in equipping students with this skill. Conversely, a link between a skill and an occupation represents that this skill is required by that occupation. For example, students are required to obtain Skills 1 and 2 to be engineers. Since students from Engineering and Computer Science majors acquire both, they can become engineers.

Following this process, the tripartite network of major-skill-occupation reduces down to the bipartite major-occupation network on the right-hand-side. The goal of the MCI is to infer the relative complexity of the skill set in each major based on the "building-block" model (left-hand-side figure) from the information contained within the flow network (right-hand-side figure). The complexity method helps to characterize the skill production structures without explicitly modeling the skill dimensionality.

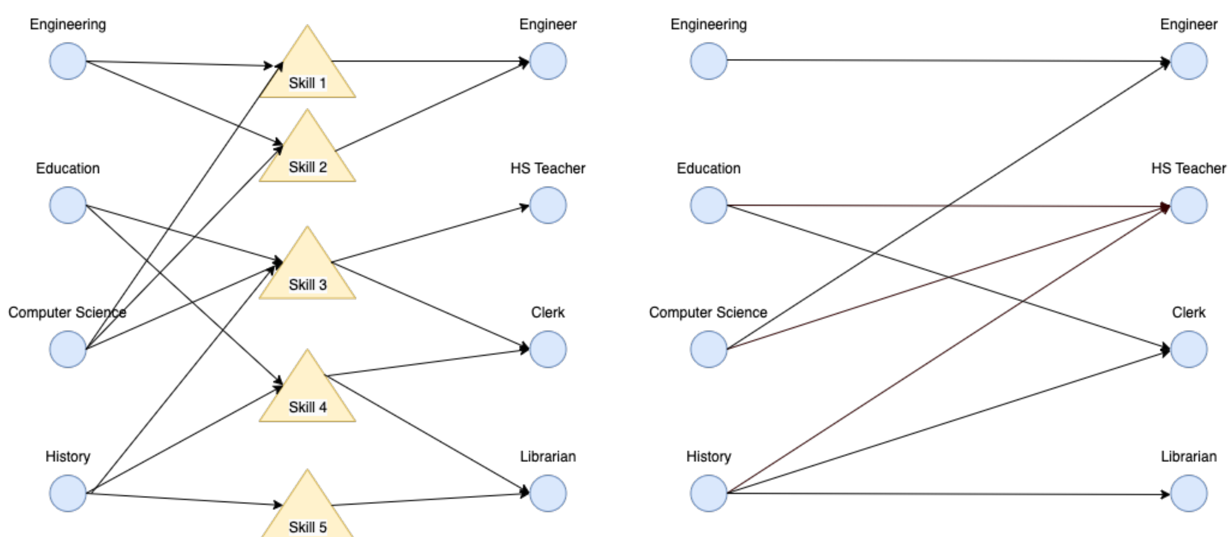

Figure A.1: Illustration of the "Building-Block" Model

Appendix B Method Details

B.1 Choice of Initial Values

When applying the method of reflection to a major-to-occupation network, it is necessary to specify an initial value of MCI for each major. We find that the specific initial values do not affect the standardized MCI following convergence. Provided that we start with an initial vector that ranks the majors according to skill specificity, the MCI invariably converges to the same value. To test this, we initiated from various points including HHI, TOP3, and GINI; in all instances, the MCI consistently converged to the same value, as evidenced in Table B.1.

Conversely, we can initiate the process from the spread count, i.e., the number of occupations associated with each major in an RCA matrix (see Section B.2). This approach precisely mirrors the original specification proposed by Hidalgo and Hausmann (2009). Here, the initial values reflect the extent to which the majors are diversified, essentially representing the inverse of specificity. Intuitively, majors that are specialized have a small spread (i.e., linked to a small number of occupations in an RCA matrix). In this case, the resulting MCI demonstrates a flipped sign version of the MCIs derived from specificity measures.

This result closely relates to the finding by Cristelli et al. (2013). In Cristelli et al. (2013), the authors assume the initial values are set by the spread count and show that the method of reflection is equivalent to finding the eigenvalues of a major-to-major similarity network, constructed from a major-to-occupation network as

$$\tilde{M}_{mm} = \sum_{j} \frac{M_{mj} M_{m'j}}{k_{m,0} \cdot k_{j,0}} ,$$

where $k_{m,0}$ denotes the number of occupations (spread count) associated with each major from an RCA matrix, and $k_{j,0}$ represents the number of majors (spread count) associated with each occupation in the same RCA network. They show that after convergence, the complexity index equates to the eigenvector of $\tilde{M}$ that corresponds to the second largest eigenvalue.[17]

Note that this relationship does not immediately apply to our application. We initialize $k_{m,0}$ by an arbitrary specificity index, not necessarily the spread count, and we even do not set any initial values for occupations, $k_{j,0}$. It is interesting that, nonetheless, we observe the MCI is insensitive to any initialization except for the sign. Clearly, there exists another mathematical nature of the Method of Reflection yet to be uncovered in this literature, but it is beyond the scope of this paper.

Table B.1: Correlation Among MCIs

	MCI_HHI	MCI_TOP3	MCI_GINI	MCI_Spread
MCI_HHI	1.0000			
MCI_TOP3	1.0000	1.0000		
MCI_GINI	1.0000	1.0000	1.0000	
MCI_Spread	-1.0000	-1.0000	-1.0000	1.0000

B.2 Network by RCA Matrix

The most important ingredient of the MCI recipe is an informative major-to-occupation network M, where M_{mj} represents a link from a major m to an occupation j. To construct this binary matrix, following the original implementation by Hidalgo and Hausmann (2009), we compute the revealed comparative advantage (RCA) of a major m for an occupation j as

$$RCA_{mj} = \frac{N_{mj}/N_m}{N_j/N},$$

where

- N_{mj} is the number of graduates from a major m who found a job in an occupation j,
- N_m is the total number of graduates from major m,
- N_j is the total number of graduates in occupation j,

[17] Note that the context of the original paper is the international trade and nodes represent countries and products. Here we keep the major/occupation notation to avoid confusion.

- N is the total number of graduates in the network.

In the binary matrix M, a link between major m and occupation j is recorded (i.e., $M_{mj} = 1$) if $RCA_{mj} > 1$. Intuitively, a major m forms a link to an occupation j if and only if the fraction of its graduates placed in j is higher than the average across all the majors.

One important caution in interpreting the Method of Reflections is the assumption that major-to-occupation mappings are primarily (if not entirely) based on the skills match. That is, for any major occupation pair, a link exists if and only if the required skills are matched. If some students, by virtue of family-resource, end up in occupations that they do not qualify for in terms of skill requirements, it is problematic to infer major values using job placements. The RCA matrix alleviates such a concern, and is notably robust against such a mismatch, by removing excess variation and focusing only on significant presences and absences (Hidalgo and Hausmann, 2009; Hidalgo, 2021; Balland et al., 2022).

This approach offers several advantages over more obvious alternatives, such as setting thresholds based on the absolute number of graduates to form a link (e.g., $M_{mj} = 1$ if $N_{mj} > 20$) or setting thresholds based on a ratio (e.g., $M_{mj} = 1$ if $\frac{N_{mj}}{N_m} > 0.05$). First, the RCA is agnostic to the population size of nodes. If a fixed count were used as a threshold, a larger major like Engineering would likely form links with every occupation, while a smaller major like Archaeology would potentially form no links at all. This approach would undoubtedly result in a loss of valuable information inherent in the original major-to-occupation flow. In contrast, the RCA allows every major to have advantages in certain occupations and disadvantages in others, as it uses relative measures. This feature makes it a more robust and insightful measure for our purposes, as it captures the nuances of the major-to-occupation network without being influenced by the sheer size of the majors.

In a related context, the RCA-based network also sidesteps the need for an additional hyperparameter related to setting a threshold. This is due to the fact that $RCA = 1$ carries a clear interpretation: it represents a comparative advantage against the average over all majors.[18] If

[18] It is possible to set the threshold to be other than 1. The result of our analysis is not sensitive to this choice.

we were to set a threshold on $\frac{N_{mj}}{N_m}$ instead of RCA, we would have to determine the value of the threshold that best extracts the information from the data. Such a threshold is likely sensitive to the aggregation method used, which adds another layer of complication. Therefore, an RCA-based network provides a more intuitive and straightforward means of analysis, thereby avoiding potential complications from arbitrary thresholds and additional hyperparameters. It thereby ensures that the approach remains consistent and replicable, regardless of the variations in the dataset or the aggregation method used.

Another approach that might be considered is to refrain from binarizing the matrix and instead utilize a weighted graph, either using RCA or raw matrix. However, as Cristelli et al. (2013) point out, it is not obvious which approach best enhances the network's informativeness in mirroring underlying activities. While a weighted matrix could theoretically encapsulate more information by preserving minor flows from a major to occupations, such a link may also be attributed to noise. Conversely, a binarized matrix concentrates on the most significant links. In other words, while a weighted graph might encompass more granularity and nuances in data, it might simultaneously introduce noise that could potentially confound the results. On the other hand, a binarized matrix simplifies the analysis and focuses on the most dominant links, potentially providing a clearer, albeit less granular, perspective on the relationships between majors and occupations.

In our analyses, we observed that the MCI demonstrates the highest explanatory power in terms of both earning and employment when constructed from a binary matrix. This was found to be superior to a weighted matrix without binarization. We also considered alternatives to RCA[19], but RCA-based MCI consistently provided the most robust explanatory power. This observation underscores that the combination of RCA and binary matrix is optimal for making the network most informative, at least in our context. Consequently, this provides fundamental justification for the use of RCA and binary matrices in our analyses.

[19] Results are available upon request.

Appendix C MCI Over Iterations

C.1 MCI Ranking Change Over Iteration

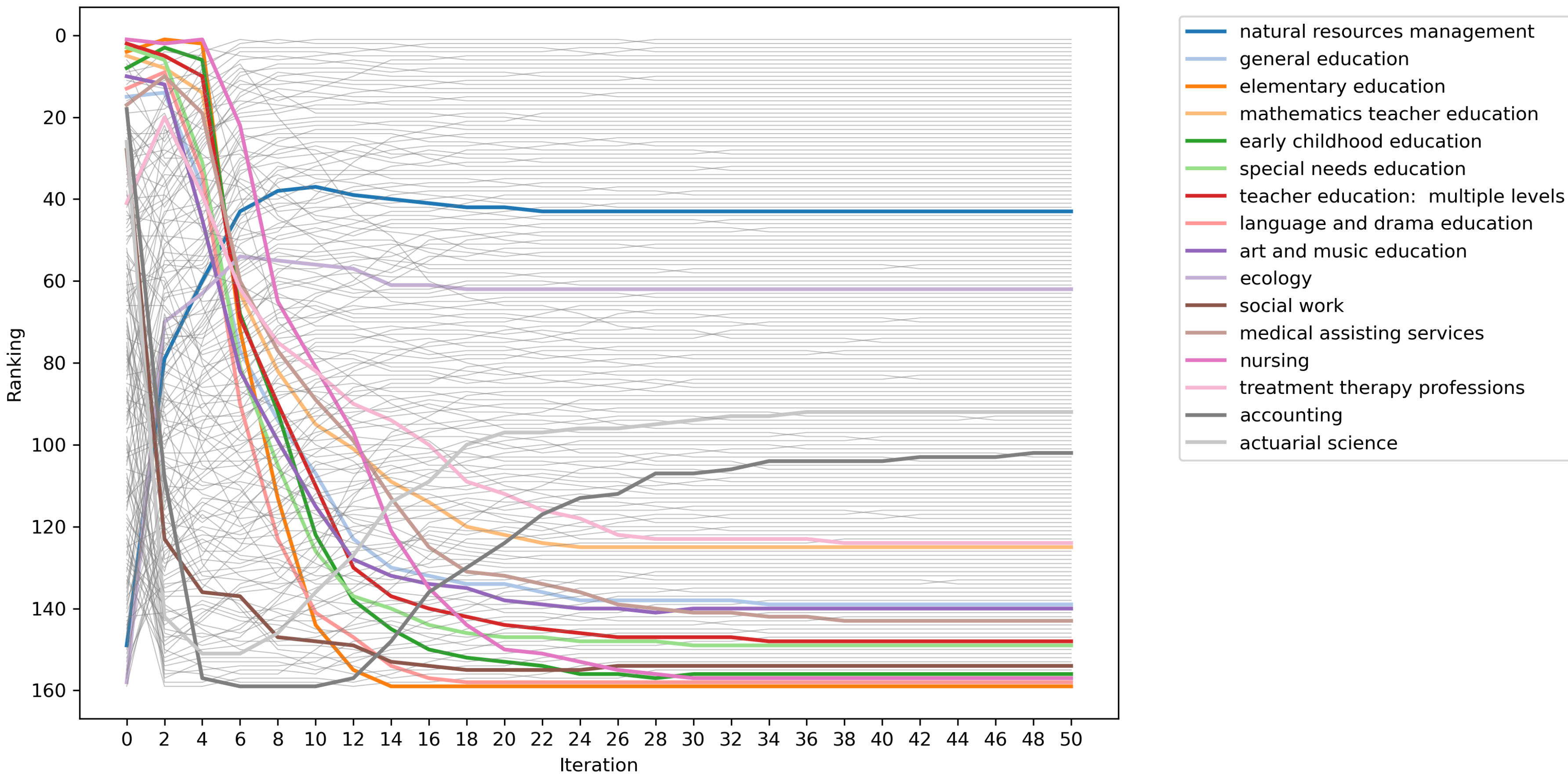

Figure C.1: MCI Ranking Change Over Iteration

C.2 Full Ranking of Majors

Table C.1: Major Ranking Over Iteration

Major Code	Major Name	HHI	Ranking HHI (Before Iteration)	Ranking MCI (After Iteration)
6107	nursing	0.7095	1	157
2312	teacher education: multiple levels	0.5417	2	148
2310	special needs education	0.5086	3	149
2304	elementary education	0.4889	4	159
2305	mathematics teacher education	0.4547	5	125
2419	petroleum engineering	0.4398	6	14
2308	science and computer teacher education	0.4123	7	54
2307	early childhood education	0.4068	8	156
2102	computer science	0.3900	9	1
2314	art and music education	0.3867	10	140
5102	nuclear, industrial radiology, and biological t...	0.3793	11	70
2309	secondary teacher education	0.3678	12	106
2313	language and drama education	0.3587	13	158
6105	medical technologies technicians	0.3517	14	75
2300	general education	0.3499	15	139
2407	computer engineering	0.3483	16	3
6104	medical assisting services	0.3427	17	143
6201	accounting	0.3201	18	102
2406	civil engineering	0.3026	19	6
2101	computer programming and data processing	0.2932	20	50
2414	mechanical engineering	0.2917	21	8
2100	computer and information systems	0.2843	22	25
2311	social science or history teacher education	0.2536	23	122
2401	aerospace engineering	0.2456	24	18
5601	construction services	0.2424	25	12
6202	actuarial science	0.2413	26	92
2105	information sciences	0.2398	27	38
5404	social work	0.2312	28	154
2405	chemical engineering	0.2306	29	5
2106	computer information management and security	0.2250	30	28
2416	mining and mineral engineering	0.2239	31	11
5201	educational psychology	0.2218	32	129
2413	materials engineering and materials science	0.2150	33	17
5008	materials science	0.2127	34	4
6108	pharmacy, pharmaceutical sciences, and administ...	0.2077	35	87
2403	architectural engineering	0.2036	36	7
2408	electrical engineering	0.2035	37	2
6004	commercial art and graphic design	0.2008	38	95
2410	environmental engineering	0.1926	39	33
2306	physical and health education teaching	0.1901	40	74
6109	treatment therapy professions	0.1833	41	124
6212	management information systems and statistics	0.1805	42	19
5901	transportation sciences and technologies	0.1580	43	35
2504	mechanical engineering related technologies	0.1562	44	29
2417	naval architecture and marine engineering	0.1551	45	9
2400	general engineering	0.1366	46	15
2399	miscellaneous education	0.1365	47	58
2418	nuclear engineering	0.1313	48	23
2412	industrial and manufacturing engineering	0.1301	49	24
2409	engineering mechanics, physics, and science	0.1291	50	10

Table C.2: Major Ranking Over Iteration

Major Code	Major Name	HHI	Ranking HHI (Before Iteration)	Ranking MCI (After Iteration)
2502	electrical engineering technology	0.1268	51	21
5002	atmospheric sciences and meteorology	0.1226	52	82
1401	architecture	0.1213	53	51
2402	biological engineering	0.1205	54	27
6209	human resources and personnel management	0.1180	55	93
6199	miscellaneous health medical professions	0.1172	56	144
4000	interdisciplinary and multi-disciplinary studie...	0.1139	57	127
2201	cosmetology services and culinary arts	0.1088	58	79
2499	miscellaneous engineering	0.1067	59	16
6207	finance	0.1055	60	100
4002	nutrition sciences	0.1047	61	118
3202	pre-law and legal studies	0.1016	62	107
2404	biomedical engineering	0.1002	63	44
2107	computer networking and telecommunications	0.0897	64	66
4901	theology and religious vocations	0.0895	65	60
3701	applied mathematics	0.0894	66	49
1302	forestry	0.0892	67	31
2500	engineering technologies	0.0888	68	22
5203	counseling psychology	0.0884	69	134
1102	agricultural economics	0.0884	70	45
3702	statistics and decision science	0.0812	71	67
6205	business economics	0.0773	72	89
5301	criminal justice and fire protection	0.0769	73	63
5004	geology and earth science	0.0766	74	53
1902	journalism	0.0759	75	153
5403	human services and community organization	0.0746	76	132
5205	industrial and organizational psychology	0.0721	77	97
4006	cognitive science and biopsychology	0.0712	78	85
2001	communication technologies	0.0712	79	76
5003	chemistry	0.0711	80	42
3605	genetics	0.0708	81	39
2503	industrial production technologies	0.0703	82	20
1100	general agriculture	0.0701	83	40
3700	mathematics	0.0694	84	37
6204	operations, logistics and e-commerce	0.0691	85	47
6211	hospitality management	0.0671	86	142
1101	agriculture production and management	0.0656	87	41
5501	economics	0.0650	88	83
2901	family and consumer sciences	0.0650	89	155
2599	miscellaneous engineering technologies	0.0641	90	13
1105	plant science and agronomy	0.0631	91	34
1103	animal sciences	0.0622	92	59
2501	engineering and industrial management	0.0620	93	26
5007	physics	0.0619	94	30
5701	electrical and mechanic repairs and technologies	0.0618	95	32
5503	criminology	0.0610	96	73
6102	communication disorders sciences and services	0.0609	97	150
6206	marketing and marketing research	0.0573	98	131
6000	fine arts	0.0564	99	119
1904	advertising and public relations	0.0562	100	138
6103	health and medical administrative services	0.0536	101	126
6099	miscellaneous fine arts	0.0530	102	78
1199	miscellaneous agriculture	0.0522	103	72
5005	geosciences	0.0521	104	48
3606	microbiology	0.0514	105	52

Table C.3: Major Ranking Over Iteration

Major Code	Major Name	HHI	Ranking HHI (Before Iteration)	Ranking MCI (After Iteration)
2602	french, german, latin and other common foreign ...	0.0514	106	136
6210	international business	0.0509	107	113
1104	food science	0.0502	108	65
6005	film, video and photographic arts	0.0493	109	104
6203	business management and administration	0.0483	110	77
5402	public policy	0.0477	111	108
3601	biochemical sciences	0.0472	112	46
4007	interdisciplinary social sciences	0.0470	113	121
6100	general medical and health services	0.0469	114	130
6106	health and medical preparatory programs	0.0461	115	81
3603	molecular biology	0.0453	116	57
5599	miscellaneous social sciences	0.0449	117	88
6299	miscellaneous business and medical administration	0.0447	118	80
6200	general business	0.0446	119	69
5505	international relations	0.0444	120	116
6007	studio arts	0.0441	121	91
6006	art history and criticism	0.0440	122	145
5507	sociology	0.0439	123	137
1901	communications	0.0437	124	147
6003	visual and performing arts	0.0430	125	152
3302	composition and speech	0.0426	126	135
3609	zoology	0.0425	127	71
5200	psychology	0.0424	128	146
3301	english language and literature	0.0421	129	151
5202	clinical psychology	0.0419	130	141
5504	geography	0.0419	131	36
1903	mass media	0.0413	132	117
5506	political science and government	0.0407	133	98
3699	miscellaneous biology	0.0407	134	55
3608	physiology	0.0405	135	99
3401	liberal arts	0.0401	136	110
5401	public administration	0.0399	137	105
6002	music	0.0398	138	103
5500	general social sciences	0.0396	139	111
6110	community and public health	0.0394	140	94
5006	oceanography	0.0387	141	56
1501	area, ethnic, and civilization studies	0.0385	142	133
6001	drama and theater arts	0.0384	143	123
4001	intercultural and international studies	0.0382	144	120
1301	environmental science	0.0378	145	64
5299	miscellaneous psychology	0.0375	146	128
3611	neuroscience	0.0371	147	86
2603	other foreign languages	0.0357	148	112
1303	natural resources management	0.0355	149	43
4101	physical fitness, parks, recreation, and leisure	0.0355	150	84
6402	history	0.0354	151	90
3402	humanities	0.0354	152	115
3600	biology	0.0353	153	68
2601	linguistics and comparative language and litera...	0.0352	154	114
6403	united states history	0.0344	155	101
4801	philosophy and religious studies	0.0335	156	109
5098	multi-disciplinary or general science	0.0332	157	61
3604	ecology	0.0328	158	62
5502	anthropology and archeology	0.0315	159	96

C.3 Regression Analysis Over Iterations

Table C.4: Earning and Employment Return on the MCI - Over Iterations

	(1)	(2)	(3)	(4)	(5)
Panel A: DV - log of Annual Income (2010 $)					
MCI_iter_2	0.0067				
	(0.0067)				
MCI_iter_4		0.0052			
		(0.0064)			
MCI_iter_10			0.0490***		
			(0.0071)		
MCI_iter_20				0.0801***	
				(0.0053)	
MCI_iter_50					0.0819***
					(0.0060)
R-squared Adj.	0.1201	0.1200	0.1253	0.1326	0.1335
Panel B: DV - Full-time Working Status					
MCI_iter_2	0.0023*				
	(0.0012)				
MCI_iter_4		0.0020*			
		(0.0012)			
MCI_iter_10			0.0103***		
			(0.0014)		
MCI_iter_20				0.0155***	
				(0.0010)	
MCI_iter_50					0.0156***
					(0.0012)
R-squared Adj.	0.0398	0.0398	0.0405	0.0412	0.0412

Note: The dependent variable in Panel A is the natural log of annual income in 2010 dollars, with 1,999,140 observations. The dependent variable in Panel B is the full-time working status with 2,591,343 observations. Gender, White, Hispanic, quadratic function of potential experience, and year-fixed effects are controlled for in all specifications. The MCIs over iterations are computed from a network connecting detailed ACS majors (which mostly correspond to CIP4) and detailed occupation categories (SOC4). Standard errors are clustered at the major-year level. Significance is as follows: one-percent=***, five-percent=**, and ten-percent=*.

Appendix D Robustness of Table 3

Table D.1: Earning and Employment Return on the MCI - Age 25-30

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: DV - log of Annual Income (2010 $)								
MCI_25-30	0.0843***				0.0931***	0.0819***	0.0834***	0.0979***
	(0.0076)				(0.0049)	(0.0063)	(0.0075)	(0.0070)
HHI		0.0257***			0.0368***			0.0520***
		(0.0050)			(0.0053)			(0.0177)
TOP3			0.0431***			0.0397***		-0.0179
			(0.0056)			(0.0063)		(0.0181)
GINI				0.0762***			0.0534**	-0.0042
				(0.0180)			(0.0214)	(0.0149)
R-squared Adj.	0.1338	0.1224	0.1252	0.1206	0.1389	0.1383	0.1341	0.1390
Panel B: DV - Full-time Working Status								
MCI_25-30	0.0154***				0.0178***	0.0152***	0.0154***	0.0212***
	(0.0015)				(0.0010)	(0.0012)	(0.0015)	(0.0014)
HHI		0.0061***			0.0086***			0.0188***
		(0.0010)			(0.0010)			(0.0033)
TOP3			0.0090***			0.0088***		-0.0111***
			(0.0012)			(0.0013)		(0.0037)
GINI				0.0026			-0.0015	-0.0167***
				(0.0044)			(0.0048)	(0.0032)
R-squared Adj.	0.0412	0.0402	0.0405	0.0398	0.0420	0.0418	0.0412	0.0422

Note: The dependent variable in Panel A is the natural log of annual income in 2010 dollars, with 1,999,140 observations. The dependent variable in Panel B is the full-time working status with 2,591,343 observations. Gender, White, Hispanic, quadratic function of potential experience, and year-fixed effects are controlled for in all specifications. The MCI and other indexes are computed from a network connecting detailed ACS majors (which mostly correspond to CIP4) and detailed occupation categories (SOC4), and important from a sample that consists of individuals between ages 25-30 instead of 25-35 as adopted in the main analysis. Standard errors are clustered at the major-year level. Significance is as follows: one-percent=***, five-percent=**, and ten-percent=*.

Table D.2: Earning and Employment Return on the MCI - Within Occupation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
DV - log of Annual Income (2010$)								
MCI	0.0452***				0.0468***	0.0412***	0.0441***	0.0390***
	(0.0027)				(0.0026)	(0.0027)	(0.0027)	(0.0033)
HHI		0.0198***			0.0217***			-0.0079
		(0.0024)			(0.0019)			(0.0079)
TOP3			0.0310***			0.0267***		0.0347***
			(0.0027)			(0.0022)		(0.0089)
GINI				0.0512***			0.0336***	0.0123
				(0.0108)			(0.0105)	(0.0114)
SOC2 FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared Adj.	0.2537	0.2515	0.2526	0.2507	0.2551	0.2553	0.2538	0.2553

Note: The dependent variable is the natural log of annual income in 2010 dollars, with 1,999,140 observations. Gender, White, Hispanic, quadratic function of potential experience, and year-fixed effects are controlled for in all specifications. The MCI and other indexes are computed from a network connecting detailed ACS majors (which mostly correspond to CIP4) and detailed occupation categories (SOC4). Standard errors are clustered at the major-year level. Significance is as follows: one-percent=***, five-percent=**, and ten-percent=*.

Table D.3: Correlation Among MCIs from Different Aggregation

	$MCI_{SOC4}^{DetailedMajor}$	$MCI_{SOC2}^{DetailedMajor}$	$MCI_{SOC4}^{CoarseMajor}$	$MCI_{SOC2}^{CoarseMajor}$
$MCI_{SOC4}^{DetailedMajor}$	1.0000			
$MCI_{SOC2}^{DetailedMajor}$	0.9046	1.0000		
$MCI_{SOC4}^{CoarseMajor}$	0.8314	0.7711	1.0000	
$MCI_{SOC2}^{CoarseMajor}$	0.8170	0.7998	0.9544	1.0000

Note: The MCIs reported in this table are computed from different networks. These networks vary based on whether the major is coded at a coarse (ACS variable *degfield* which corresponds to CIP2) or detailed level (ACS variable *degfieldd* which mostly corresponds to CIP4) and whether the occupation is measured at a coarse (SOC2) or detailed level (SOC4).

Table D.4: Earning and Employment Return on the MCI - Controlling for Major Categories and Occupation

	(1)	(2)	(3)	(4)	(5)
DV - log of Annual Income (2010$)					
MCI	0.0452***		0.0448***		0.0204***
	(0.0027)		(0.0038)		(0.0046)
STEM		0.0736***	0.0012		
		(0.0075)	(0.0087)		
Arts and Humanities				-0.0221**	-0.0070
				(0.0107)	(0.0107)
Business				0.0929***	0.0952***
				(0.0110)	(0.0107)
Education				-0.0600***	-0.0390***
				(0.0100)	(0.0098)
Engineering				0.2132***	0.1808***
				(0.0110)	(0.0137)
Professional				0.1036***	0.1176***
				(0.0116)	(0.0124)
Science				0.0621***	0.0428***
				(0.0119)	(0.0108)
Social Science				0.0589***	0.0710***
				(0.0152)	(0.0144)
SOC2 FE	Yes	Yes	Yes	Yes	Yes
R-squared Adj.	0.2537	0.2519	0.2537	0.2584	0.2587

Note: The dependent variable is the natural log of annual income in 2010 dollars, with 1,999,140 observations. Gender, White, Hispanic, quadratic function of potential experience, and year-fixed effects are controlled for in all specifications. The MCI is computed from a network connecting detailed ACS majors (which mostly correspond to CIP4) and detailed occupation categories (SOC4). Standard errors are clustered at the major-year level. Significance is as follows: one-percent=***, five-percent=**, and ten-percent=*.

Appendix E Major Complexity Index Rankings Over Time

Table E.1: MCI Ranking Over Time

Major Code	Major Name	2009-2013	2015-2019
2408	electrical engineering	1	4
2407	computer engineering	2	2
2419	petroleum engineering	3	11
2414	mechanical engineering	4	13
2102	computer science	5	3
2403	architectural engineering	6	35
2413	materials engineering and materials science	7	18
2100	computer and information systems	8	15
2406	civil engineering	9	1
2409	engineering mechanics, physics, and science	10	29
2412	industrial and manufacturing engineering	11	41
2400	general engineering	12	8
2502	electrical engineering technology	13	5
5601	construction services	14	9
2401	aerospace engineering	15	26
2503	industrial production technologies	16	27
2599	miscellaneous engineering technologies	17	16
2504	mechanical engineering related technologies	18	30
2418	nuclear engineering	19	7
2500	engineering technologies	20	14
2405	chemical engineering	21	6
2417	naval architecture and marine engineering	22	10
2499	miscellaneous engineering	23	22
2501	engineering and industrial management	24	21
5008	materials science	25	23
5007	physics	26	17
1105	plant science and agronomy	27	37
2410	environmental engineering	28	32
5701	electrical and mechanic repairs and technologies	29	19
1100	general agriculture	30	44
1302	forestry	31	38
6212	management information systems and statistics	32	28
2105	information sciences	33	12
1101	agriculture production and management	34	48
2402	biological engineering	35	31
2106	computer information management and security	36	24
2416	mining and mineral engineering	37	25
5901	transportation sciences and technologies	38	20
1303	natural resources management	39	42
3606	microbiology	40	67
1401	architecture	41	33
6204	operations, logistics and e-commerce	42	55
5004	geology and earth science	43	45
2399	miscellaneous education	44	51
1102	agricultural economics	45	63
5005	geosciences	46	46
3701	applied mathematics	47	36
2107	computer networking and telecommunications	48	73
5003	chemistry	49	43
3601	biochemical sciences	50	52

Table E.2: MCI Ranking Over Time

Major Code	Major Name	2009-2013	2015-2019
5504	geography	51	34
2101	computer programming and data processing	52	40
1103	animal sciences	53	70
3700	mathematics	54	53
5006	oceanography	55	75
5301	criminal justice and fire protection	56	47
3702	statistics and decision science	57	65
2308	science and computer teacher education	58	78
6200	general business	59	76
3600	biology	60	66
6299	miscellaneous business and medical administration	61	84
2404	biomedical engineering	62	39
3699	miscellaneous biology	63	59
3605	genetics	64	58
6203	business management and administration	65	80
1301	environmental science	66	50
5002	atmospheric sciences and meteorology	67	77
1199	miscellaneous agriculture	68	92
3604	ecology	69	68
5098	multi-disciplinary or general science	70	57
5102	nuclear, industrial radiology, and biological t...	71	56
1104	food science	72	61
6004	commercial art and graphic design	73	114
3603	molecular biology	74	54
5501	economics	75	85
6205	business economics	76	94
2201	cosmetology services and culinary arts	77	89
3609	zoology	78	74
6201	accounting	79	101
2001	communication technologies	80	64
2306	physical and health education teaching	81	82
2309	secondary teacher education	82	147
6207	finance	83	118
4101	physical fitness, parks, recreation, and leisure	84	79
4901	theology and religious vocations	85	49
5503	criminology	86	72
6210	international business	87	102
6209	human resources and personnel management	88	87
6402	history	89	88
6007	studio arts	90	81
6110	community and public health	91	98
4801	philosophy and religious studies	92	106
6105	medical technologies technicians	93	60
6099	miscellaneous fine arts	94	71
2603	other foreign languages	95	115
5500	general social sciences	96	107
5401	public administration	97	91
2601	linguistics and comparative language and litera...	98	123
5402	public policy	99	127
4006	cognitive science and biopsychology	100	119
3611	neuroscience	101	86
5502	anthropology and archeology	102	110
6106	health and medical preparatory programs	103	90
6202	actuarial science	104	69
6000	fine arts	105	108

Table E.3: MCI Ranking Over Time

Major Code	Major Name	2009-13	2015-2019
6002	music	106	103
6211	hospitality management	107	141
5506	political science and government	108	96
4001	intercultural and international studies	109	122
4007	interdisciplinary social sciences	110	124
1903	mass media	111	121
6108	pharmacy, pharmaceutical sciences, and administ...	112	62
6001	drama and theater arts	113	129
6005	film, video and photographic arts	114	83
3608	physiology	115	97
3401	liberal arts	116	93
5599	miscellaneous social sciences	117	111
6103	health and medical administrative services	118	130
6100	general medical and health services	119	117
2311	social science or history teacher education	120	100
6206	marketing and marketing research	121	151
4002	nutrition sciences	122	112
5505	international relations	123	116
5299	miscellaneous psychology	124	134
5205	industrial and organizational psychology	125	125
3302	composition and speech	126	136
3402	humanities	127	104
6403	united states history	128	99
2300	general education	129	135
6199	miscellaneous health medical professions	130	113
1904	advertising and public relations	131	146
2602	french, german, latin and other common foreign ...	132	126
1501	area, ethnic, and civilization studies	133	131
6003	visual and performing arts	134	137
3301	english language and literature	135	148
3202	pre-law and legal studies	136	105
1901	communications	137	154
4000	interdisciplinary and multi-disciplinary studie...	138	120
2312	teacher education: multiple levels	139	138
6006	art history and criticism	140	143
5507	sociology	141	133
5403	human services and community organization	142	132
6109	treatment therapy professions	143	95
2305	mathematics teacher education	144	109
5200	psychology	145	139
5202	clinical psychology	146	152
2310	special needs education	147	158
5203	counseling psychology	148	153
1902	journalism	149	150
2313	language and drama education	150	145
6104	medical assisting services	151	142
5201	educational psychology	152	149
2314	art and music education	153	128
2307	early childhood education	154	157
6102	communication disorders sciences and services	155	140
5404	social work	156	144
2901	family and consumer sciences	157	155
6107	nursing	158	156
2304	elementary education	159	159

Appendix F NSSE Data Descriptive and Selection Exercise 1

To check the robustness of our main result in Table 3 and further understand the Major Complexity Indices, we exploit pooled data from the National Survey of Student Engagement (NSSE) for the years 2005-2011.[20] Our final NSSE sample contains 93,193 full-time seniors, who are between the ages of 18 and 23, and with no double major.[21] NSSE contains rich student-level data on the types of assignments and tasks performed across majors and information on students' backgrounds.

We compute the major average of SAT scores, parental education, and percent of international students for each major as well as major characteristics surveyed from students, such as knowledge and skills developed through college education and hours spent on coursework at the CIP2 level, and merge with our ACS sample. In total, we are able to match 22 majors (out of 22 majors in NSSE; out of 38 majors in ACS) using CIP2 between the two datasets.[22,23]

Table F.1 below shows the first selection exercise using major-level ability and preference proxies. The results are similar to our main results. As shown in Panel A, the income return to MCI is 10.15% with basic controls in this sample, and it decreases to 8.25% when major average SAT scores, parental education levels, and percent of international students are controlled for. The employment return in Panel B is also similar across the two columns.

[20]The confidential data from the NSSE contains no individual institution or student identifier and can be obtained by application. See `https://nsse.indiana.edu/nsse/index.html` for more details.

[21]Two small majors, Library/archival science (4 students) and Undecided (17 students), are dropped.

[22]The CIP2 majors that are in ACS but not NSSE include 10) communications technologies/technicians and support services. 12) personal and culinary services. 19) family and consumer sciences/human sciences. 22) legal professions and studies. 25) library science. 29) military technologies. 31) parks, recreation, leisure, and fitness studies. 39) theology and religious vocations. 41) science technologies/technicians. 43) security and protective services. 44) public administration and social service professions. 46) construction trades. 47) mechanic and repair technologies/technicians. 48) precision production. 49) transportation and materials moving. 51) health professions and related clinical sciences.

[23]If we instead merge at the CIP4 level, we are able to match 48 majors (out of 60 majors in NSSE; out of 159 majors in ACS) using CIP4 between the two datasets.

Table F.1: Earning and Employment Return on the MCI - Controlling Major-Level Features from NSSE

	(1)	(2)
Panel A: DV - log of Annual Income (2010 $)		
MCI	0.1015***	0.0825***
	(0.0039)	(0.0045)
SAT_V		0.0033***
		(0.0004)
SAT_M		0.0012***
		(0.0003)
InternationalStudent		1.0409***
		(0.3568)
Edu_Father		-2.3485***
		(0.2617)
Edu_Mother		-0.0659
		(0.2010)
R-squared Adj.	0.1460	0.1507
Panel B: DV - Full-time Working Status		
MCI	0.0195***	0.0188***
	(0.0008)	(0.0008)
SAT_V		-0.0004***
		(0.0001)
SAT_M		0.0010***
		(0.0001)
InternationalStudent		-0.7826***
		(0.0778)
Edu_Father		-0.4873***
		(0.0438)
Edu_Mother		0.0522
		(0.0371)
R-squared Adj.	0.0448	0.0472

Note: The dependent variable in Panel A is the natural log of annual income, with 1,697,541 observations. The dependent variable in Panel B is the full-time working status with 2,198,729 observations. Gender, White, Hispanic, quadratic function of potential experience, and year-fixed effects are controlled in all specifications. The MCI is computed from a network connecting detailed-level ACS majors (which mostly correspond to CIP4) and detailed occupation categories (SOC4). Standard errors are clustered at the major-year level. Significance is as follows: one-percent=***, five-percent=**, and ten-percent=*.

Appendix G NLSY97 Data Descriptive and Selection Exercise 2

To further check the robustness of our main result in Table 3, we merge the MCI computed from the ACS, constructed from the coarse major (variable *degfield*) and detailed occupation (SOC4), with the National Longitudinal Survey of Youth 1997 (NLSY97), which collects extensive information on individuals' educational experience and labor market outcomes, particularly concerning preexisting abilities and academic qualifications.[24]

Following the approach of Light and Schreiner (2019), we identify each individual's college major by cycling through the Post-Secondary Transcript Study (PTRAN) degree variables. These variables are recorded as two-digit College Course Map (CCM) codes across all institutions associated with an individual. When degree variables are missing or recorded as "liberal arts and sciences, general studies and humanities" (CCM code=24), which applies to approximately 62% of the sample, we then cycle through the PTRAN field of study variables, which capture information from different transcript sections. In the end, we successfully identify the college major for 891 individuals. These individuals only have bachelor's degrees, have no missing information about their college major, and do not have double majors from any institutions. We manage to match 36 majors (out of 37 majors in NLSY97; out of 38 majors in ACS) using CIP2 between ACS and NLSY97 datasets.[25] It's worth noting that there are 8 majors each containing fewer than 5 individuals. Due to the small sample size (as explained below), we included all matched majors.

Our analysis begins with 891 individuals, yielding 16,929 individual-by-year observations, all of whom have valid college major information. Restricting the data to observations of individuals older than 25 years, who have already obtained a bachelor's degree within the years, and who are not currently enrolled in school, we are left with 880 individuals, accounting for 6,160 individual-by-year observations. For income analysis, as demonstrated in Table G.1 Panel A, our sample reduces to 371 individuals (representing 2,252 observations), all of whom are working

[24]The public-use data from the NLSY97 can be downloaded at `https://www.nlsinfo.org/investigator/pages/login`.

[25]The CIP2 majors present in ACS but absent in NLSY97 include 25) library science and 48) precision production. Conversely, the CIP2 majors present in NLSY97 but absent in ACS include 32) basic skills.

full-time, earning at least \$500 annually, and have no missing information on the covariates. For the employment analysis (Table G.1 Panel B), the sample consists of 390 individuals (representing 3,086 observations).

Although our sample size is significantly smaller, the results align with our main analysis. A one standard deviation increase in the MCI results in approximately a 7-8% increase in salary and enhances the probability of full-time employment by approximately 5%. These estimates are adjusted based on ability measures (SAT/ACT, AFQT score, HS GPA) and household information (average household income during ages 15-19, parental education). These adjustments are very similar to the estimates with basic controls presented in column (1). Using more detailed proxies for ability, such as SAT Verbal and Math scores, the 10 components of the ASVAB test scores, and HS GPA in specific fields including English, foreign language, math, social science, and life science, does not significantly alter the results. However, it does result in a much smaller sample size due to the absence of observations.

Table G.1: Earning and Employment Return on the MCI - Controlling Individual Features from NLSY

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: DV - log of Annual Income (2010 $)						
MCI	0.0859***	0.0781***	0.0767***	0.0806***	0.0843***	0.0753***
	(0.0175)	(0.0174)	(0.0175)	(0.0170)	(0.0175)	(0.0175)
SAT_ACT		0.0702***				0.0107
		(0.0155)				(0.0213)
AFQT			0.0033***			0.0015**
			(0.0005)			(0.0007)
HS_GPA_Overall				0.1590***		0.1136***
				(0.0321)		(0.0395)
logHH_Income (Age 15-19)					0.0504**	0.0509***
					(0.0198)	(0.0194)
Edu_Father					0.0169***	0.0117**
					(0.0057)	(0.0057)
Edu_Mother					-0.0020	-0.0003
					(0.0052)	(0.0051)
R-squared Adj.	0.1924	0.1995	0.2021	0.2010	0.2014	0.2113
Panel B: DV - Full-time Working Status						
MCI	0.0505***	0.0536***	0.0520***	0.0498***	0.0510***	0.0523***
	(0.0084)	(0.0083)	(0.0082)	(0.0083)	(0.0083)	(0.0080)
SAT_ACT		-0.0273***				-0.0284**
		(0.0101)				(0.0142)
AFQT			-0.0006*			-0.0000
			(0.0003)			(0.0005)
HS_GPA_Overall				0.0145		0.0446**
				(0.0185)		(0.0218)
logHH_Income (Age 15-19)					-0.0161**	-0.0138*
					(0.0072)	(0.0074)
Edu_Father					-0.0140***	-0.0134***
					(0.0043)	(0.0042)
Edu_Mother					0.0057	0.0069*
					(0.0037)	(0.0036)
R-squared Adj.	0.0427	0.0450	0.0432	0.0426	0.0492	0.0509

Note: The dependent variable in Panel A is the natural log of annual income, with 2,252 observations. The dependent variable in Panel B is the full-time working status with 3,086 observations. Gender, White, Hispanic, quadratic function of potential experience, and year-fixed effects are controlled in all specifications. The MCI is computed from a network connecting coarse-level ACS majors (which correspond to CIP2) and detailed occupation categories (SOC4). Standard errors are clustered at the major-year level. Significance is as follows: one-percent=***, five-percent=**, and ten-percent=*. Using detailed ability proxies such as SAT_V and SAT_M; 10 components of ASVAB test scores; HS GPA in detailed fields including English, foreign language, math, social science, and life science, do not significantly change the results, although it results in a much smaller sample without missing observations.