Major Complexity Index and College Skill Production [*†‡]

Xiaoxiao Li

Sebastian Linde

Hajime Shimao

August 20, 2022

Abstract

We propose an easily computable measure called the Major Complexity Index (MCI) that captures the latent skills taught in different majors. By applying the Method of Reflections to the major-to-occupation network, we construct a scalar measure of the relative complexity of majors. Our measure provides strong explanatory power of major average earnings and employment. Further evidence suggests that the MCI is strongly associated with advanced skills such as quantitative problem-solving, and the use of computing technology. We also provide a two-stage algorithm to partial out selection on observables which opens up possibilities of applying the complexity measure in various contexts.

Keywords: Major-to-Occupation Network; Skill Acquisition; Method of Reflections; Complexity.
JEL classifications: I2, J2

---

[*]We are especially indebted to Timothy Bond, Caroline Hoxby, Michael Lovenheim, and Kevin Stange for suggestions and advice that helped us improve the manuscript substantially. We are also grateful to Joseph Altonji, 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, Lois Miller, Kevin Mumford, Philip Oreopoulos, Steve Ross, Jeffrey Smith, Yang Song, Jones Todd, Evan Totty, Sarah Turner, Muzhe Yang, and participants at Eastern Economic Association Annual Conference, Western Economic Association International, 2021 CSWEP CeMENT Mentoring Workshop, Midwest Economics Association Annual Conference, NBER Economics of Education Spring 2021 Program Meeting, and the Society of Labor Economists for helpful comments.

[†]The authors declare no relevant or material financial interests that relate to the research described in this paper. This paper uses public use data files that are available for download at the National Science Foundation (see: `https://ncsesdata.nsf.gov/datadownload/`) for the main analysis. Secondary analysis uses confidential data from the National Survey of Student Engagement (NSSE), and no individual institutions or students are identified in the paper. This data can be obtained by application. (see: `https://nsse.indiana.edu/nsse/index.html`).

[‡]Contact information: Xiaoxiao Li, Villanova University, xiaoxiao.s.li@villanova.edu; Sebastian Linde, Medical College of Wisconsin, slinde@mcw.edu; Hajime Shimao, Santa Fe Institute, hshimao@santafe.edu.

1 Introduction

The return to education varies widely across fields of study in college (Altonji et al., 2012, 2016).[1] This heterogeneity is partly due to the fact that different majors send students to differing sets of occupations. For example, students with a petroleum engineering degree can find high-paying jobs as petroleum engineers, which are not easily accessible to students from many other majors. From the perspective of human capital accumulation, this linkage between college majors and students' occupational outcomes largely reflects the match of multi-dimensional skills. That is, through various college majors, students acquire different sets of skills, which are subsequently employed in diverse job tasks.

A natural question to ask, then, would be: which major equips students with the most applicable set of skills? There is considerable difficulty when striving to answer this question. First of all, the skill acquisition process through different college majors is unobservable and thus needs to be indirectly inferred. Moreover, it is challenging and potentially dangerous for researchers to define what constitutes a "skill". In the previous literature, skills tend to be defined from our intuitions, possibly in low dimensions, such as verbal and quantitative, or cognitive and non-cognitive (e.g., Heckman et al., 2006; Kinsler and Pavan, 2015).

In this paper, we take a drastically different approach. By adopting the so-called "building-block" model proposed by Hidalgo and Hausmann (2009) in the context of international trade, we create an indirect measure of the skill set acquired through college education for each major. The intuition of the "building-block" approach is rather straightforward. In the context of education, each skill can be viewed as a different type of building block, like a Lego piece of a particular shape. Each occupation is a Lego model, which requires a unique combination of building blocks. In order for a college graduate to find a job within an occupation, she needs to obtain the required set of building blocks to form such a model. Now, the role of college majors is clear in this analogy: Each major is a bucket of building blocks from which students can pick up the required pieces.

[1] Altonji et al. (2012) shows that the wage gap between electrical engineering and general education majors is almost as large as that between college and high school graduates.

What we need is a measure that captures the number and variety of blocks available in each bucket (i.e. college major). One obvious candidate is the number of occupations a major sends its students to. For instance, petroleum engineering and education administration majors both send graduates to eight distinct occupations, which may seem to imply that they teach similarly complex skills. However, this simple counting method would miss important information on the difficulty and specificity of skills that can be acquired from different majors. For example, one major may send students to occupations that only require a minimal skill set students could acquire from a large number of majors. In contrast, students from another major may find jobs in occupations where students from other majors cannot easily get in. In such cases, we would think the latter equips students with a set of more valuable, non-substitutable skills.

To construct such a measure, we adopt the "Method of Reflections" technique introduced by Hidalgo and Hausmann (2009).[2] This method exploits the rich information embedded within a major-to-occupation flow network and recovers the latent structure of necessary building blocks by an iterative algorithm. This paper is, to our knowledge, the first attempt to apply the Method of Reflections in the education and labor context.

Specifically, by incorporating information from the entire bipartite major-to-occupation network, we are able to uncover the underlying tripartite network connecting college majors to the skills they produce, and occupations to the skills they require. We call this measure the Major Complexity Index (MCI). The MCI takes into account relevant neighboring information in the network such that majors with exactly the same spread (number of occupations a major sends students to) can yield very different complexity ranks. Using the example above, petroleum engineering and education administration majors are both linked to eight occupations: the former is tied with occupations that are rarely accessible to other majors (less "ubiquitous" in the terminology of Hidalgo and Hausmann, 2009), and consequently it is connected to other complex majors, which altogether returns a high major complexity index rank (2/137); while the latter maps students to

[2]Hidalgo and Hausmann (2009) introduce this method to take the bipartite network between countries and their exported goods, and measure the latent production capacities and technologies that countries possess in order to produce the basket of observed exports. After its first introduction, this Economics Complexity Index (ECI) has been extensively utilized in the international trade literature (e.g. Tacchella et al., 2012; Mealy et al., 2019).

occupations that are linked to many other majors, each of which is not highly ranked in terms of major complexity, and in turn yields a lower MCI rank (135/137). 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.

We explore three versions of the MCI according to a: (1) Binary flow matrix; (2) Weighted flow matrix based on the distribution of students within a major-to-occupation network; and (3) Controlled flow matrix based on the average marginal effects from a multinomial choice model that accounts for potential bias due to selection on observables. One caveat of using the Method of Reflections in education-labor contexts is selection bias. Our modified two-stage algorithm (Controlled MCI) allows us to reduce selection on observables, and as such, it opens up possibilities of using the complexity measure in settings where selection bias would otherwise be a concern.

Our empirical analysis employs individual-level information, including college major choice and occupational outcome, from the National Survey of College Graduates (NSCG) data in 2003, 2010, and 2015, to construct a bipartite major-to-occupation network, one for each year, and examine the relationship between the MCI measures and average earnings as well as employment differentials across college majors. Our results indicate that the MCI reveals important aspects of college majors that matter to earnings (especially in recent years) and employment of college graduates. For example, using NSCG 2015 data, a one standard deviation increase in the Binary MCI raises salary by $\$12,821$ or 16.9%, and boosts employment by 1.78 percentage points, and the R^2 increases substantially when the MCIs are introduced. An interesting observation within our study is that the power of the MCI to explain across-major wage differentials has increased considerably between the early 2000s and 2015, and decreased for employment. As we track the MCI over time, we observe that some majors experience considerable changes in the ranking, which may reflect structural changes in the labor market. For instance, Actuarial Science was ranked 96th in 2003, but rose to 37th in 2010, and further to 24th in 2015.

In order to better understand what the MCI captures, we combine our major-to-occupation flow data from the NSCG with major level characteristics from the National Survey of Student Engagement (NSSE). Our results suggest that high MCI majors tend to have students with better pre-college academic qualifications, e.g. higher SAT scores in all three dimensions (mathematics, verbal reasoning, and writing ability), especially through a strong positive correlation with SAT math. In addition, high MCI majors are more intense and demanding in terms of time spent preparing for class, completing problem sets, and working on longer written assignments. More interestingly, in terms of knowledge and skills acquired through college education, students within high MCI majors tend to report further development of quantitative and practical problem solving, and the use of computing and information technology; but not in terms of basic skills such as written and spoken communication (which presumably are developed mainly through primary and secondary education), or in acquiring job-related knowledge and skills.

Overall, our work contributes to a rich literature on the 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 extremely difficult to measure, particularly along the dimensions that are of interest by employers. On the occupation side, there are studies documenting the skills required by different occupations (see, e.g., Graetz and Michaels, 2018). However, it is not an easy task to quantify the same skills acquired in college majors. For instance, how do we measure the programming skills that students can obtain from an economics major on average? Secondly, skills are presumably high dimensional. In a seminal paper, Cunha and Heckman (2007) identify two dimensions of skills, cognitive and non-cognitive, by means of a factor model.[3] These two dimensions may likely represent the most important distinction of skills, especially for early childhood education. In the context of college education, however, more fine-grained skill categories are necessary. For instance, leadership, a specific skill within the non-cognitive domain, is documented to have predictive power of potential earnings (Kuhn and Weinberger, 2005). A recent paper by Deming (2017) highlights the importance of social skills in the labor market. Thirdly, another important yet under-explored aspect is the complementarity

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

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 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.

In summary, all of these challenges make our approach particularly appealing. Our proposed method does not intend to solve these issues, but rather circumvent them. To see this, it is important to note that the "building-block" model fully captures the high-dimensional and combinatorial nature of skills, while the computation of the MCI avoids explicit estimation of the acquired skills and directly uncover the relative value of majors by exploiting the match of students between college majors and occupations, which is logically analogous to the revealed preference approach (Avery et al., 2013) but in terms of revealed skills rather than preferences. There are surprisingly few easily-computable quantitative descriptions of college majors, with limited examples such as major average SAT scores. Our comprehensive measure of major complexity is simple to compute with a minimum data requirement, and importantly, does not rely on student-reported data on major features. Such a method can complement the existing structure-based approaches (see, e.g., Altonji et al. 2012), and facilitates us to better understand the unobserved skill production process through college majors. Furthermore, the MCI provides a convenient and informative reference for both prospective students in choosing college majors, and education administrators in strategic planning, especially in resource-constrained environments.

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

2 Methods

Suppose we have $\mathscr{M}$ college majors and $\mathscr{O}$ occupations. Let N be a $\mathscr{M} \times \mathscr{O}$ flow matrix that represents the majors to occupations network. We first consider the flow matrix to be a binary matrix where $N_{m,o} = 1$ if and only if there exists at least one student who graduates from major m and finds a job in occupation o, and $N_{m,o} = 0$ otherwise. Using the Method of Reflections introduced by Hidalgo and Hausmann (2009), we iteratively calculate the value for each major and occupation according to equation (1) and (2), respectively:

$$k_{m,b} = \frac{1}{k_{m,0}} \sum_{o \in \mathscr{O}} N_{m,o} k_{o,b-1} \tag{1}$$

$$k_{o,b} = \frac{1}{k_{o,0}} \sum_{m \in \mathscr{M}} N_{m,o} k_{m,b-1} \tag{2}$$

where $b = 1, ..., B$ refers to the number of iteration and the initial values ($b = 0$) are the raw counts from the flow matrix:

$$k_{m,0} = \sum_{o \in \mathscr{O}} N_{m,o}$$

$$k_{o,0} = \sum_{m \in \mathscr{M}} N_{m,o}$$

Intuitively, $k_{m,0}$ represents the "spread" of major m in terms of the total number of occupations students can get in after graduating from major m. Similarly, $k_{o,0}$ represents the "specificity" of occupation o in terms of the count of majors that can place students in occupation o.[4] On the major side, a larger spread indicates higher complexity (as such a major can lead students to more occupations), while on the occupation side, smaller specificity implies higher complexity (as such an occupation accepts students only from limited majors). Upon convergence, the even iteration on the major side produces what we call the Binary Major-Complexity-Index (MCI).[5] The following example illustrates the Method of Reflections in the context of major-to-occupation flow network.

[4]The term "spread of a major" and "specificity of an occupation" correspond to "diversification of a country" and "ubiquity of a product" in Hidalgo and Hausmann (2009). Either way, they represent the number of links directly connected to a node.

[5]A similar index can be computed on the occupation side which is beyond the focus of this paper. See Section 5 for further discussion.

Example of Binary MCI

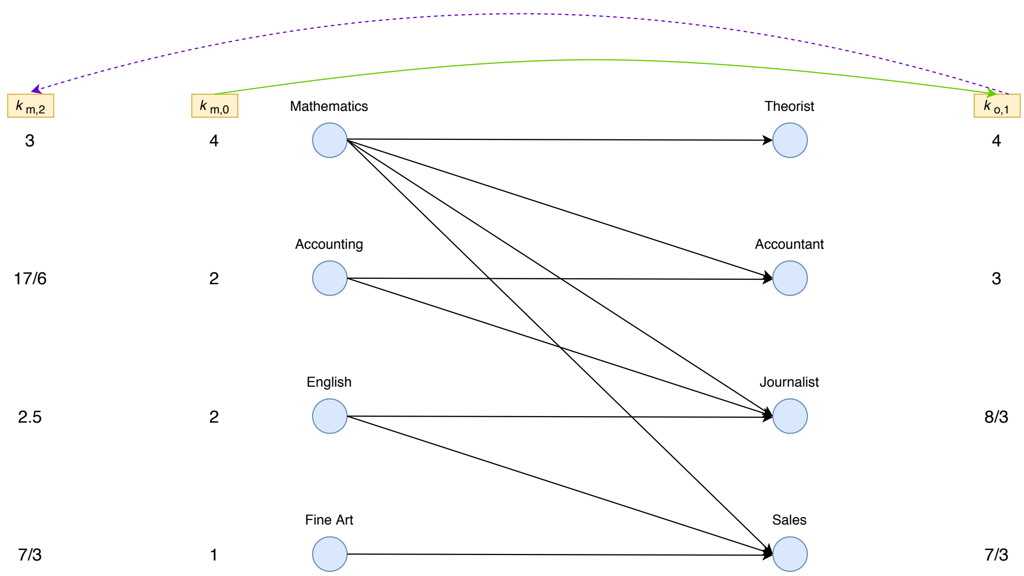

Figure 1: Bipartite Major-to-Occupation Network Example

Suppose there are four majors and four occupations, linked in a network as shown in Figure 1. For $m = Mathematics$,

$$k_{m,0} = 4,$$

indicating that students with a math degree find jobs in four distinct occupations (i.e. the spread of Mathematics is four). This is the naive complexity index at $b = 0$ where Mathematics is ranked in first place, followed by Accounting, English, and lastly Fine Art. The intuition is that the more occupations a major directs students to, the more complex the major is in terms of skills taught. However, this naive measure does not account for the accessibility of occupations. Using the examples in Figure 1, the spread of Accounting and English majors are both two, with a common occupation outlet (Journalist). The difference is that Accounting can lead students to become Accountants, which is less accessible to students from other majors, when compared to Sales that is the alternative occupational outcome to an English major.

To incorporate this useful information within the network, we exploit the iterative procedure called the Method of Reflections. We first use the spread $k_{m,0}$ to calculate the complexity of each occupation, $k_{o,1}$, in the first iteration ($b = 1$). As an example, for $o = Accountant$,

$$k_{o,1} = \frac{1}{k_{o,0}} \sum_{m \in \mathscr{M}} N_{m,o} k_{m,0}$$
$$= \frac{1}{2}(1 \times 4 + 1 \times 2 + 0 \times 2 + 0 \times 1) = \frac{4+2}{2} = 3.$$

That is, there are two majors, Mathematics and Accounting, that place students as Accountants ($k_{o,0} = 2$), and the average spread ($k_{m,0}$) of these two majors is calculated to be 3. The same calculation can be carried out for all occupations $o \in \mathscr{O}$, as shown in Figure 1, $k_{o,1}$ column. Again, using the examples in this network, Journalist and Sales are both connected to three majors (the occupation specificity $k_{o,0} = 3$), but the spread of majors that are linked to Journalist is, on average, greater than that for Sales. As a consequence, Journalist is ranked higher than Sales in this iteration. Importantly, majors are connected to one another through the occupation nodes within a network, and this intermediate step $k_{o,1}$ is essential to reflect such connections.

We then iterate back ($b = 2$) to update the complexity measure of each major, $k_{m,2}$, using those values obtained in $k_{o,1}$. See the solid and dashed arrows on the top of Figure 1 for a graphic illustration of the iteration procedure. Back to the example of $m = Mathematics$,

$$k_{m,2} = \frac{1}{k_{m,0}} \sum_{o \in \mathscr{O}} N_{m,o} k_{o,1}$$
$$= \frac{1}{4}\left(4 + 3 + \frac{8}{3} + \frac{7}{3}\right) = 3,$$

where $N_{m,o} = 1$ for all $o \in \mathscr{O}$ since the Mathematics major is linked to all four occupations, and thus is omitted from the expression above. Intuitively, the average score of $k_{o,1}$ is 3 for Mathematics. Recall, $k_{o,1}$ captures the average spread of majors that are connected to an occupation. Following this logic, $k_{m,2}$ is basically the average of average spread of majors that are connected to one another through common occupational outcomes. What's embedded in this averaging process is the adjustment of major spread by occupation specificity. As shown in Figure 1, the complexity index of Accounting is updated from $k_{m,0} = 2$ to $k_{m,2} = 17/6 = 2.8\dot{3}$, while the adjustment for the

English major is only from $k_{m,0} = 2$ to $k_{m,2} = 2.5$ in this iteration. As a result, Accounting is now ranked higher than the English major, precisely for the aforementioned reason that Accounting can send students to occupations that are difficult to get into (i.e. Accountant), and in turn it is connected to other majors that are relatively more complex in terms of skills taught on average.

While we stop the illustration at $b = 2$, we iterate on this procedure until the ranking exhibits full convergence. The complexity index on the occupation side, $k_{o,b}$ 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 $k_{o,b}$ 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 Binary Major-Complexity-Index (MCI) on the major side, which is a specificity-adjusted spread 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. We elaborate on how a tripartite network of major-skill-occupation reduces down to a bipartite major-occupation network in Appendix A. The goal of the MCI is to infer the relative complexity of the skill set in each major based on the "building-block" model (a tripartite network) from the information contained within a bipartite flow network.[6]

It is helpful to see how our framework relates to the important notion of general versus specific human capital (e.g. Kinsler and Pavan, 2015). As the skill-portfolio analysis literature (e.g. Silos and Smith, 2015) suggests, some skills are very specific to certain occupations while others are more generally applicable. Conversely, some skills are easily attainable from many majors (e.g. oral communication) while others are taught in specific majors (e.g. understanding of thermody-

[6]See Hidalgo and Hausmann (2009) and Hidalgo (2021) for more extensive discussions.

namics that is commonly taught in a petroleum engineering major). The iteration procedure, based on the "building-block" model, favors majors that teach skills that are specific to majors, but not necessarily specific to occupations. To understand this, let's consider the following four cases: (1) Suppose a general skill, such as verbal or written communication that is required by the vast majority of occupations, is part of the training in most majors, then this skill does not contribute to differentiate majors. (2) Similarly, if there is any skill that is specific to an occupation but is widely taught across majors (although such a case is probably rare), then this skill does not help to rank majors. (3) On the contrary, if a major teaches a general skill, say data analytic or management that is increasingly valued by various occupations, yet not taught by many majors, then this major would be relatively highly ranked within our framework. (4) Lastly, it's often the case that, when a major prepares students with a set of knowledge and skill that are specific to certain occupations (e.g. understanding of thermodynamics), such a major tends to be highly ranked. The key takeaway from this exercise is that a complex major tends to teach skills that are valuable in the marketplace yet not trained in many other majors. And through those skills developed, such a major can direct students to occupations that are not easily accessible, and in turn is connected to other majors that are more complex in terms of skills taught, both of which are crucial to compute a high MCI.

Weighted MCI

The MCI calculation above utilizes a binary adjacency matrix. One could also construct the index using a weighted adjacency matrix instead. That is, if 100 students with a math degree find jobs in Theorist (10), Accountant (40), Journalist (20), and Sales (30), we then use the percentage of students from major m to each occupation as the proper weights (e.g., 0.1, 0.4, 0.2, 0.3) rather than equal weights (e.g. $\frac{1}{4}$) to calculate $k_{m,b}$. Similarly, if 100 students end up being Accountants, with 40 from Math and 60 from Accounting, we use the percentage of students from each major to occupation o as the proper weights (e.g., 0.4, 0.6) rather than equal weights (e.g. $\frac{1}{2}$) to calculate $k_{o,b}$. In this case, we call the converged even-iteration value on the major side the "Weighted MCI".

Controlled MCI

The Weighted MCI exploits useful information contained in the distribution of students within a major-to-occupation network. One caveat, however, especially in education and labor context, is potential selection bias. That is, the apparent link between major m and occupation o could be partially due to factors other than skills match, such as gender preference. To account for selection on observable characteristics, we construct what we call the "Controlled MCI". Below we provide a general recipe, followed by our simplest implementation as an illustration.

In the first step, we estimate the choice probability of each occupation conditional on student characteristics, together with college major information. That is, for each student i, let (m_i, o_i, x_i) be a tuple of his or her major choice, occupation choice, and other characteristics, such as age and gender.[7] We then estimate a multinomial choice model that yields an estimate of the conditional probability, $Prob(o_i = o|m_i, x_i)$, for each occupation $o \in \mathscr{O}$. In the second step, we use these estimated choice probabilities of occupations to construct the controlled major weight matrix such that the (m, o)-th element represents the average marginal effect of graduating with major m on the likelihood of being in occupation o.[8] Similarly, we can construct the controlled occupation weight matrix by estimating the conditional probability, $Prob(m_i = m|o_i, x_i)$, for each major $m \in \mathscr{M}$, and assigning the average marginal effect of being in occupation o on the likelihood of coming from major m as the (o, m)-th element in the matrix. Intuitively, these average partial effects represent, on average, how likely a person is in an occupation from each major, and vice versa.

As long as the data permits, the choice probabilities can be estimated using any complicated structural model (see, e.g., Altonji et al. 2012). In this paper, for illustration purposes, we adopt a

[7]Due to data limitation, we are only able to control for gender and age in the individual regressions described below using fully disaggregated major categories (i.e. CIP6). To identify the model, it is required that variables have sufficient variation in every major and every occupation. For instance, Geological Engineers occupation in 2003 only contain males (12 students) and Pre-school Teacher Education major in 2015 only contain females (42 students), and therefore are dropped from the analysis. For the same reason, we exclude race and other controls as they would result in omission of many more majors and occupations in our dataset. Using more aggregated major categories (i.e. CIP2), we also control for race and parental education, and the results are very similar. We discuss the major aggregation results further in section 4.2.

[8]The Method of Reflections requires all entries of the matrix to be non-negative. Thus we normalize the matrix by the min-max scaling.

simple multinomial logit model. That is, for each occupation $o \in \mathscr{O}$, we estimate the model:

$$Prob(o_i = o|m_i, x_i) = \Lambda\left(\sum_{m \in \mathscr{M}} \alpha_{m,o} \mathscr{I}(m_i = m) + \beta_o x_i\right),$$

where $\mathscr{I}(.)$ is an indicator function, and for each major $m \in \mathscr{M}$, we estimate:

$$Prob(m_i = m|o_i, x_i) = \Lambda\left(\sum_{o \in \mathscr{O}} \alpha_{o,m} \mathscr{I}(o_i = o) + \beta_m x_i\right).$$

Note that the purpose of any multinomial choice model here is to construct the weight matrices that capture the joint distribution of majors and occupations as before while controlling for observables. Once we obtain the weight matrices, we can implement the Method of Reflections to construct the Controlled MCI. This is done exactly as in the Weighted MCI approach, however, we now replace $N_{m,o}$ in equation (1) with the (m,o)-th element of the controlled major weight matrix, and $N_{m,o}$ in equation (2) with the (o,m)-th element of the controlled occupation weight matrix.

Discussion of Methods

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.[9] That is, for any major-occupation pair, a link exists if and only if the required skills are matched. Presumably, there may exist two types of mismatches between majors and occupations in any realized network: (1) a major produces sufficient skills to be linked with an occupation, yet a link is not observed; (2) a major is matched with an occupation even though it does not teach the necessary skill set.

The former concern is alleviated if the sample is sufficiently large and students have heterogeneous preferences. In other words, it is actually desirable to have some students who learned sufficient skills from a major, but chose less skill-demanding occupations due to factors such as gender preference (e.g. students from an engineering major find jobs as an elementary school teacher). Consequently, such an observed link indicates that this major equips sufficient skills to

[9]This critical assumption is acknowledged in the original Hidalgo and Hausmann (2009) study as well.

direct students to less skill-demanding occupations, although some of those skills are potentially wasted. It is worth noting that the Binary MCI is more robust against this type of mismatch than the Weighted MCI. To see this, consider an engineering major where selection is possibly prominent (e.g. low female ratio, high SAT math score). So long as one student chooses to be in an occupation like elementary school teacher, a binary link is recorded and utilized to compute the Binary MCI. In contrast, the proportion of students in such an occupation, which is subsequently used to construct the Weighted MCI, is small and negligible. Put differently, the weighted MCI is more fragile to selections that may result in the first type of mismatch between majors and occupations.

If the second type of mismatch is salient in a network, it presents a different concerning issue to the methodology. For instance, a student may end up in an occupation by virtue of family-resource that he or she does not qualify in terms of skill requirements, which makes it problematic to infer major complexity, particularly with the Binary MCI, if he or she is the only individual that connects this major-occupation pair. One possible remedy to this issue is to impose an arbitrary threshold when constructing a major-occupation network (e.g. at least 10 students from a major to an occupation in order for a linkage to be recorded). However, by doing so, the coverage of possible occupational outlets for a given major necessarily shrinks, which then give rise to the first type of mismatch. Following a similar rationale, the Weighted MCI is more robust against the second type of mismatch than the Binary MCI, since it is less influenced by any spurious linkages. Taken together, there is a trade-off between the two types of mismatches, and it is therefore beneficial to take advantages of both the Binary and Weighted MCI as references.

The Controlled MCI is an improved version of the Weighted MCI where the distribution of students within a major-to-occupation network is adjusted to be more reflective of the skills match. However, it is critical to point out that the interpretability of this approach (within our basic implementation in this paper) hinges on a selection on observables assumption, which is evidently restrictive. Much of the recent literature on returns to college majors has focused on identifying the causal impact of majors on future earnings, allowing for selection on unobservables across majors. It follows that, to fully isolate the complexity of the latent skills associated with any given major, one needs either a rich set of controls or more assumptions (i.e. a more complicated structural model). Hence, our Controlled MCI results presented below should be interpreted with caution. Our purpose here is to introduce this general procedure so that policymakers with much better data availability (such as SAT scores, high school GPA, family background and financial resources, etc.) can compute a robust major complexity index as a convenient and informative reference that reflects the skills taught in different majors and that are required by different occupations.

3 Data

Our main data is sourced from the National Survey of College Graduates (NSCG) administered by the National Science Foundation. We use three cross-sectional datasets from the 2003, 2010, and 2015 surveys for our main analysis. One convenience of the NSCG data is that it provides individual-level information on schooling (i.e. major choice) along with occupational history, both of which are critical to construct a bipartite major-to-occupation network. We create three major-to-occupation networks, one for each survey year, with the following two data restrictions: First, in order to concentrate our analysis on the occupational placement of more recent graduates, we restrict our dataset to contain only individuals below the age of 40. Second, in order to minimize noise from spurious major-to-occupation linkages, we only keep majors and occupations with a minimum of at least 5 individuals in them.[10,11]

Table 1 provides descriptive statistics for our main outcome variables—major average salary in 2015 dollars (Salary) and major average employment rate (Employment Rate)—along with the major spread (Spread) and Major Complexity Index (MCI) obtained using the binary (MCI_B),

[10]After imposing these restrictions, we have 27,852 observations in 2003, 24,315 observations in 2010, and 38,685 observations in 2015, that are used to construct major-to-occupation matrices. We are able to connect 140 majors to 92 occupations in 2003, 135 majors to 100 occupations in 2010, and 137 majors to 100 occupations in 2015. The number of majors and occupations are not restricted to be the same across years since they potentially reflect structural changes in the labor market. For instance, as technology reshaped the landscape in the 2000s, the national survey in 2010 and 2015 included trending occupations such as Computer Network Architect, Computer Programmer, Software Developers, Web Developer, etc. that were not captured in the 2003 survey.

[11]We also conduct a robustness exercise where we further restrict the sample to individuals without any graduate degrees nor occupational licenses. Using the confined sample with only 14,519 observations in 2015, the results (available upon request) are quantitatively similar to those presented in Table 2 and 3, Panel A.

weighted (MCI_W), and controlled (MCI_C) approaches. As shown in Panel A, there are a total of 137 majors in 2015 with an average major-level salary of $66,758 and employment rate of 91 percentage points. The mean major spread is about 37, implying that, on average, majors direct students to 37 distinct occupations. The major complexity indices are standardized to have within-sample mean of zero and variance of one.[12] Using the Binary MCI in 2015 as an illustration, the General Business major has a standardized index of zero, while Physics is one standard deviation above, and Elementary Teacher Education is one standard deviation below. Panel B and C display the major-level information of 2010 and 2003, respectively. The mean salary is slightly higher in 2010 compared to 2015; however, the variance, as well as the range (min and max), are smaller during the recovery period of the 2008 financial crisis. The lower mean employment rate tells a similar story. In contrast, the mild recession in 2001 does not have the same long lasting effects. The mean salary and employment rate are both higher in 2003 compared to 2010, where 76% of the majors, including Medical preparatory programs, Economics, and Accounting, experienced higher payments in real terms back in 2003, while other majors, such as Counseling Psychology, Statistics, and Petroleum Engineering enjoyed a wage premium surge in the late 2000s. Lastly, it's worth noting that the major spread is lower in 2010 compared to 2015, and even smaller in 2003.

To further understand what the Major Complexity Indices capture, we combine the 2015 NSCG data with a pooled cross-sectional dataset from the National Survey of Student Engagement (NSSE) for the years 2010-2011.[13] The latter contains rich student level data on the types of tasks and assignments performed across majors, such as knowledge and skills developed through college education as well as hours spent on homework and papers. It also contains pre-college information such as SAT scores. We are able to map 78 majors in the NSSE to 70 majors in the NSCG.[14] See Appendix D, Table D.1 for the summary statistics of the NSSE variables used in the matched dataset.

[12]The indices are obtained after 250 iterations to ensure full convergence.

[13]NSSE surveys both freshmen and seniors in college. Our final NSSE sample of 2010 and 2011 data include 43% freshmen and 57% seniors who are most likely in the labor market by the time of 2015.

[14]There are cases where multiple NSSE majors are mapped to the same NSCG major. For instance, Civil engineering and Urban planning (NSSE) are both linked to Civil engineering (NSCG).

4 Results

The empirical results are organized into three subsections. We elaborate on the intuition of the MCI in Section 4.1, and present major-level regression results for both salary and employment in Section 4.2. Lastly, Section 4.3 provides an in-depth decomposition analysis of the MCI measure.

4.1 Spread versus MCI

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 illustrates how the ranking of college majors change over iterations ($k_{m,0}$ to $k_{m,10}$) using the NSCG 2015 dataset. Here, we highlight two majors: Petroleum Engineering (yellow) and Education Administration (red). While both majors have the same spread ($k_{m,0}$) as their students are mapped into the same number of occupations (8 in both cases), the type of occupations that their students end up in differ considerably, and in turn we find that their major complexity indices ($k_{m,b}$ where $b =$ even number after convergence) are different. In the case of the Petroleum Engineering major, students are mapped into occupations that are rarely accessible to other majors such as petroleum and chemical engineers, and consequently is connected to other complex majors like aerospace, aeronautical and astronautical engineering on average. 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 (2/137). The reverse is true for the Education Administration major in which students are mapped into easily accessible occupations such as secondary school teaching and educational administration work, which in turn yields a lower MCI rank (135/137). The important takeaway here is that majors with exactly the same spread can yield very different complexity rankings due to the specific skill combination gained through each major which results in distinct occupational outcomes.

4.2 Major Mean Wage and Employment Rate Analysis

While one could hypothesize that majors with greater spreads offer students a larger choice set of occupations and therefore generate higher option value, this does not appear to be the case as shown in Figure 3. This figure fits a linear regression between: (i) major mean salary against the spread ($k_{m,0}$) in the left-hand-side plot; and (ii) major mean salary against the Binary MCI in the right-hand-side plot. There is in fact no relationship between average salary and the spread, while there is a strong positive correlation between the average salary and the MCI. This highlights the intuition that variety of occupations by itself does not contain much information regarding the important skills acquired in each major. To discuss the value of a major, we need to consider questions such as: Do the jobs available from this major require a complex skill combination that is not easily accessible from other majors, or do they demand only a minimal skill set that anyone could have? What comparative advantages does a major prepare its students in terms of skills and knowledge taught that are valuable and non-replaceable in the labor market?

The ordinary least squares results in Table 2 confirm our intuition above. Major spread is neither statistically nor economically significant in either level or log salary regressions, while the MCI measures are statistically significant at 1 percent level in explaining the earning differentials across college majors and substantially increase the explanatory power. Using NSCG 2015 data in Panel A, a one standard deviation increase in the Binary MCI raises major mean salary by $\$12,821$ in column (2) or 16.9% in column (6), and the effect is smaller using the Weighted MCI: $\$11,472$ in column (3) or 15.5% in column (7). The Controlled MCI paint a similar picture as shown in column (4) and (8) with even smaller estimates when we adjust the major-to-occupation flow matrix on the basis of age and gender, as described in Section 2. A one standard deviation increase in the Controlled MCI raises salary by $\$10,739$ or 14.3%. The R^2 increases from 0.001 in column (1) to 0.522 in column (2) when the Binary MCI is controlled for, and it is 0.426 in column (3) and

0.374 in column (4) when the Weighted and Controlled MCI is employed, respectively. Here we note another interesting observation that the explanatory power of the MCI increases over iterations when we apply the Method of Reflections. As shown in Appendix B, the R^2 increases from 0.321 to 0.519 between the 2nd and 10th iteration, which is surprisingly large considering the Method of Reflections is merely a simple manipulation of the same data. Intuitively, the complexity of a major emerges from the number of marketable skills, the depth of each skill, and the interactions among those skills. And these skill sets provided by majors are predictive of students' potential earning outcomes. Similar patterns can be observed using NSCG 2010 and 2003 data in Panel B and C, respectively. Interestingly, the estimates are fairly similar in 2003 and 2010, but notably smaller in magnitude compared to the ones in 2015. The R^2 is also smaller in 2003 and 2010.

Furthermore, the major complexity index is also statistically significant at the 1 percent level in explaining the employment rate differences across college majors, and it considerably adds the explanatory power, as shown in Table 3. Using NSCG 2015 data in Panel A, a one standard deviation increase in the Binary MCI raises employment rate by 1.78 percentage points in column (2), or 1.97 percentage points in column (3) and 1.86 percentage points in column (4), using the Weighted and Controlled MCI, respectively. Interestingly, the estimated effect is larger in 2010 and the largest in 2003, and the R^2 is the highest in 2003. It would be interesting to explore plausible explanations of these observed chronological patterns. Nonetheless, a thorough examination is beyond the scope of this paper. In Appendix C, we report the MCI rankings over time, which may reflect structural changes in the labor market. For instance, Actuarial Science was ranked 96th in 2003, but rose to 37th in 2010, and further to 24th in 2015.

In the appendices, we also explore three robustness checks on the results in Table 2 and 3. First, as shown in Table D.2, we examine the robustness of the MCI in explaining wage and employment differentials while controlling for major average SAT scores and other student-reported major characteristics. Even with a smaller sample size in the matched dataset, after controlling for students' academic qualifications (i.e. removing potential positive selection bias on preexisting abilities), majors with higher complexity scores still produce substantially higher average earnings.

And in terms of the employment margin, controlling for additional major features results in larger estimates of the return to major complexity. We note in Appendix D that, since the development of advanced knowledge and skills is the central channel through which the MCI affects earning and employment outcomes, controlling for additional major characteristics may be an over-control for our purpose.

Second, while it is debatable what the "right" level of major aggregation is, we conduct our main analysis at CIP6 level since it typically corresponds to students' course-taking activities and major declarations. Nevertheless, to check the sensitivity of the method to different levels of major aggregation, we report the major rankings using CIP4 and CIP2 in Tables C.4-5 and Table C.6, respectively, which appear to be qualitatively consistent with the major ranking based on CIP6 as shown in Tables C.1-3 (i.e. engineering and hard science majors are relatively highly ranked). There are, however, important heterogeneities within the aggregated major groups. For instance, Education major is ranked 17/30 using CIP2. As we zoom into CIP6, it is interesting to see that Mathematics teacher education is highly ranked (59/137) compared to Education administration (135/137). Moreover, Appendix E presents results of mean wage and employment using more aggregated major categories for both CIP4 and CIP2, and these results are very similar to those presented in Table 2 and 3 (Panel A). We conduct another robustness check where additional controls are added using CIP2, and the results are very similar. For instance, a one standard deviation increase in the Controlled MCI raises salary by $\$6,855$ or 9.9% when age and gender are controlled for, and the estimates decrease slightly to $\$6,660$ or 9.7% when additional variables (race and parental education) are added. See the notes under Table E.1-2 for more details.

Third, Appendix F investigates the robustness of the MCI in predicting labor market outcomes compared to traditional approaches. The traditional approaches classify majors by rough area of study (such as STEM versus non-STEM; Art and Humanities, Social Science, etc.). Several interesting features are noteworthy: First, as shown in Panel A column (2), earnings are on average substantially higher for STEM majors, but the effect disappears after controlling for the MCI in column (4), and the R^2 increases remarkably from 0.142 to 0.526. Second, even after introducing the macro fields of study (9 categories where Agriculture is the omitted group), majors with higher complexity scores still present considerably higher average earnings as shown in Panel A, columns (5) and (6). Third, we do not observe similar patterns in employment outcome, at least using the 2015 NSCG data, as shown in Panel B.

4.3 Major Complexity Index Decomposition

Our preceding analyses suggest that the MCI reveals important aspects of college majors that matter to earnings (especially in recent years) and employment of college graduates. In order to better understand what the MCI captures, we combine the major-to-occupation flow data from the NSCG 2015 with major level characteristics from the National Survey of Student Engagement (NSSE) for the years 2010-2011. In total, we are able to match 78 majors in the NSSE to 70 majors in the NSCG.

Table 4 provides the pairwise correlation between the Binary MCI 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.

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 quantitative problem solving, and the use of computing and information technology are strongly positively correlated with the MCI measure, with an estimated correlation of about 0.57 and 0.52, respectively. 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 a strong negative 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 within prior schooling), but who primarily report developing their ability to think critically and analytically, as well as the ability to analyze and solve quantitative and practical problems. Another observation worth noting is that, surprisingly, the MCI measure does not appear to be correlated with acquiring job or work-related knowledge and skills. Presumably, the robustness of the MCI in explaining wage and employment rate differentials across college majors is due to quantitative and analytical skills captured by the MCI rather than direct knowledge regarding the job content.

In terms of time spent and efforts, we find that high MCI majors tend to have students who on average report spending longer hours preparing for classes, completing problem sets, and working on longer written assignments. If the time spent on studying can be viewed as a proxy for coursework intensity and difficulty, then this suggests that high MCI majors are more demanding on students and require them to invest more efforts into their schooling, which could generate higher payoffs in the future.

We also conduct a multivariate analysis to further evaluate how the MCI is related with major-specific characteristics, and the results are reported in Table D.3.[15] Importantly, as shown in column (1), there is still about half of the variation in the MCI left once we control for sorting into majors by the SAT scores (adjusted $R^2 = 0.494$). Holding the SAT scores fixed, additional major-specific variables, such as knowledge and skills developed in column (4) and time spent in column (5), are predictive of the MCI, but to a limited extent (adjusted $R^2 = 0.633$ and 0.609, respectively). Even with all variables controlled for in column (6), the adjusted R^2 is still below 0.65, which suggests that the MCI is not simply a repackaging of traditional measures. The major ranking based on the MCI is also not just as simple as traditional categorizations such as STEM vs non-STEM majors, or broad fields of study (e.g. Art and Humanities, Social Science, etc.), although engineering and hard science majors tend to be highly ranked. See Appendix C, Tables C.1-3 for more details on the major complexity ranking.

There are surprisingly few easily-computable quantitative descriptions of college majors. Our comprehensive measure of major complexity is simple to compute with a minimum data require-

[15]Some of our intuition above from the pairwise correlation remain consistent in the linear regressions, although, some skills are necessarily entangled (e.g. the correlation between analyzing quantitative problems and using computing and information technology is 0.727), and therefore changing one dimension while holding others constant may be an unrealistic exercise. See Appendix D for more detailed analysis.

ment, and it provides strong explanatory power in understanding average earning and employment variations across college majors. Most importantly, it does not rely on student-reported data on major features. In our view, it is fruitful to exploit major-to-occupation flow networks which can complement the existing structure-based approaches (see, e.g., Altonji et al. 2012), and facilitates us to better understand the unobserved skill production process through college majors.

5 Discussion

How are advanced skills formed through college education via different majors, and is this skill production process responding to the changing nature of the economy? 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, 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 the major-to-occupation flow network and construct a scalar measure of the relative complexity in terms of skills taught in different majors. Our measure of complexity provides strong explanatory power in understanding average earning and employment variations across college majors. An interesting observation within our study is that the power of the MCI to explain across-major wage differentials has increased considerably between the early 2000s and 2015. One possible reason for this may lay with quantitative skills enjoying an increasingly large wage premium in the labor market. Recent work by Acemoglu and Autor (2011) suggests that the rapid diffusion of new technologies could distort the earning distribution in a way that benefits high-skilled workers. Our additional results provide further support along this avenue and suggest that the MCI strongly relates to advanced skills such as quantitative and practical problem solving, and the use of computing and information technology.

The major complexity indices exhibit rankings of college majors[16] 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. Furthermore, as technology exponentially advances, skills valued by the labor market constantly evolve (Deming, 2017; Graetz and Michaels, 2018). The time trend of the major complexity ranking intuitively displays how each major is changing its relative position by modifying the bucket of "building blocks" as a response to the market demand for skills.

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 a budget-constrained circumstance. For instance, 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 are there-through 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 ranking of major complexity to better understand the comparative advantages of their own institution, and build the short- and long-term strategic plans accordingly.

[16]See Appendix C, Tables C.1-3 for the major ranking based on the Binary MCI at CIP6 level; and similarly Tables C.4-5 and Table C.6 for the major rankings using CIP4 and CIP2, respectively.

One important caution in using the Method of Reflections is the assumption that major-to-occupation mapping is primarily (if not entirely) based on the skills match. That is, the underlying model is that college majors produce skills, different occupations require different skill combinations, and a graduate finds a job if and only if the skills are matched. Conversely, if a linkage does not exist between a major and an occupation, it is only because the skills are not sufficient rather than nobody wants that job. This concern is ameliorated to some extent by the fact that the National Survey of College Graduates (NSCG) surveys a large sample of college graduates in every wave, with likely heterogeneous preferences. However, the potential selection bias remains as an important caveat in applying this method to the education-labor environment. For instance, the apparent match between a major and an occupation could be based on factors other than skills, such as family resources or gender preference. Another contribution of this paper is to provide a two-stage algorithm (Controlled MCI) to partial out the selection on observables. Our algorithm opens up the possibilities of employing the complexity measure in various other contexts.

Despite the noted limitations, the Method of Reflections and the complexity index are widely adopted in the international trade and development literature (See Hidalgo 2021 for a summary of applications in those fields) 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, our paper only demonstrates the MCI dynamics over time. It certainly would be very interesting to explore spatial variations in major complexity. For instance, the MCI of a given major may vary across different types of institutions, characterized by important attributes such as public vs. private sector, college quality, geographic location, etc. (Clotfelter 1999; Dale and Krueger 2002, 2014; Black and Smith 2004; Hoxby 2009; Ehrenberg 2012; Dillon and Smith 2020; among others). Note, in order to compute meaningful major complexity indices, one would need sufficient observations in each major-occupation pair within each type of institution. Second, we only observe and consider labor market outcomes at a particular time point in life. Needless to say, investment in higher education can create benefits for potentially an entire career (Hoxby, 2020). It is a promising future extension to investigate whether the MCI explains various lifetime earnings risks (Ryoo and Rosen, 2004; Dillon, 2018). Third, in this paper, we focus on 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 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.

Avery, C. N., M. E. Glickman, C. M. Hoxby, and A. Metrick (2013). A revealed preference ranking of us colleges and universities. *The Quarterly Journal of Economics 128*(1), 425–467.

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.

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

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.

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

Dillon, E. W. (2018). Risk and return trade-offs in lifetime earnings. *Journal of Labor Economics 36*(4), 981–1021.

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*.

Graetz, G. and G. Michaels (2018). Robots at work. *Review of Economics and Statistics 100*(5), 753–768.

Heckman, J. J., J. Stixrud, and S. Urzua (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor economics 24*(3), 411–482.

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.

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

Mealy, P., J. D. Farmer, and A. Teytelboym (2019). Interpreting economic complexity. *Science advances 5*(1), eaau1705.

Ryoo, J. and S. Rosen (2004). The engineering labor market. *Journal of political economy 112*(S1), S110–S140.

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.

Tacchella, A., M. Cristelli, G. Caldarelli, A. Gabrielli, and L. Pietronero (2012). A new metrics for countries' fitness and products' complexity. *Scientific reports 2*, 723.

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

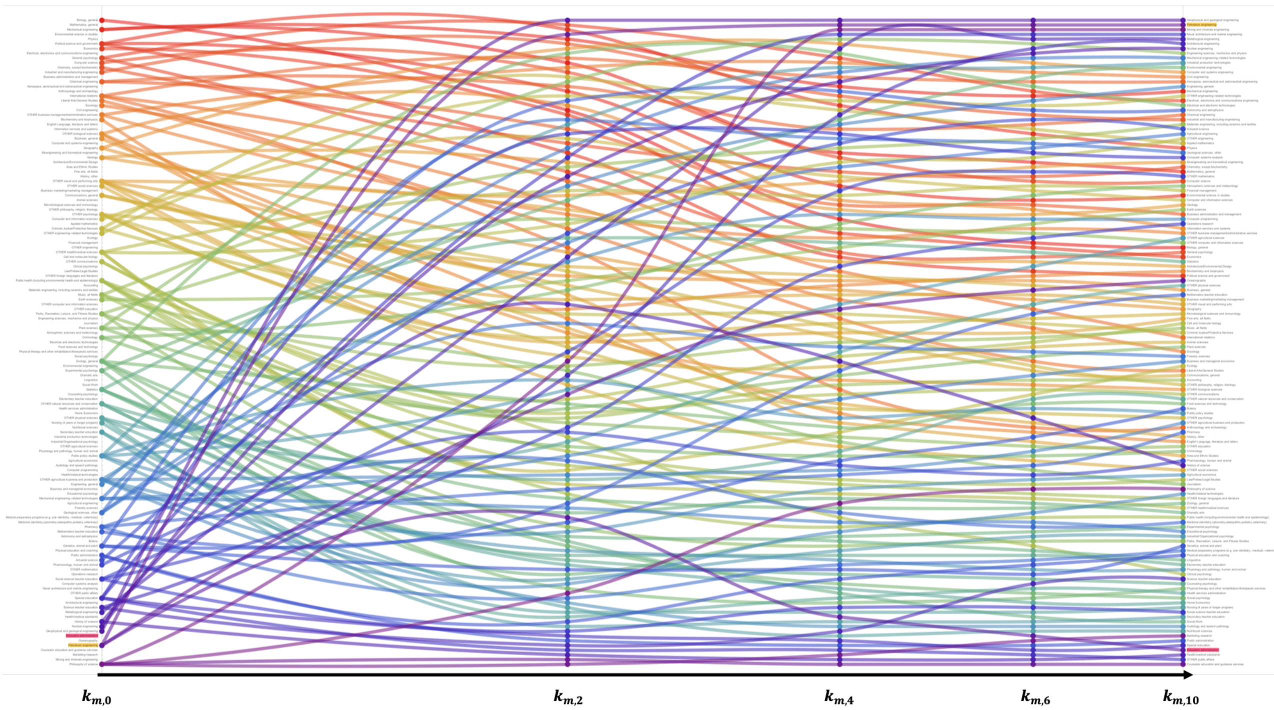

Figure 2: MCI ranking over iterations (Binary Adjacency Matrix Using NSCG 2015 Data). Iterations: 0,2,4,6, and 10 are reported. Two majors (Petroleum Engineering and Education Administration) that both map into 8 occupations are highlighted as yellow and red, respectively. Majors that are higher-up in the plot have higher MCI.

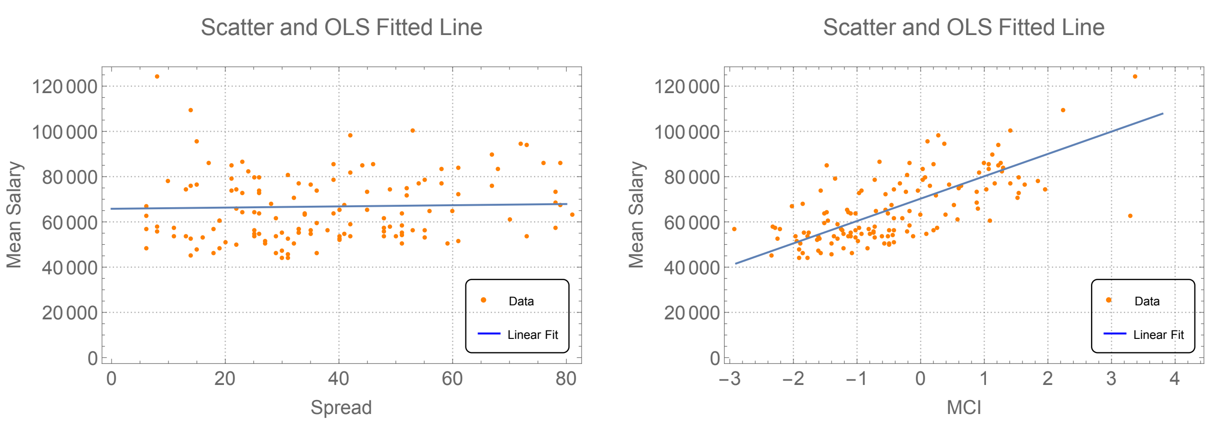

Figure 3: Scatter and Ordinary Least Squares Regression Lines. Using 2015 NSCG data, the left-hand-side (LHS) plot explores the relationship between major specific mean-salary and major-spread; while the right-hand-side (RHS) plot explores the relationship between major specific mean-salary and the Binary MCI.

Table 1: Summary statistics

	Mean	Std. Dev.	Min	Max	N
Panel A: NSCG 2015 Data					
Salary (2015$)	66,758.08	17,575.34	43,839.43	152,092.29	137
Employment Rate	91.01	4.62	75	100	137
Spread	36.69	19.19	6	81	137
MCI_B	0.00	1.00	-2.02	3.646	137
MCI_W	0.00	1.00	-1.78	3.286	137
MCI_C	0.00	1.00	-1.471	4.217	137
Panel B: NSCG 2010 Data					
Salary (2015$)	67,337.70	15,857.12	40,398.04	128,075.88	135
Employment Rate	88.36	5.41	69.23	100	135
Spread	31.01	17.16	5	81	135
MCI_B	0.00	1.00	-1.866	3.79	135
MCI_W	0.00	1.00	-2.001	3.406	135
MCI_C	0.00	1.00	-1.566	4.663	135
Panel C: NSCG 2003 Data					
Salary (2015$)	74,632.16	17,789.34	41,513.88	160,896.52	140
Employment Rate	89.14	5.52	66.67	100	140
Spread	29.96	16.19	3	71	140
MCI_B	0.00	1.00	-2.104	3.329	140
MCI_W	0.00	1.00	-2.425	3.858	140
MCI_C	0.00	1.00	-1.723	5.042	140

Note: Panel A reports the summary statistics for the NSCG 2015 data and major average statistics are computed using data from 38,685 students; Panel B for the NSCG 2010 data and major average statistics are computed using data from 24,315 students; and Panel C for the NSCG 2003 data and major average statistics are computed using data from 27,852 students. MCI measures are standardized to have sample mean 0 and standard deviation 1.

Table 2: Major Level Wage Regressions

	(1) Salary	(2) Salary	(3) Salary	(4) Salary	(5) ln(Salary)	(6) ln(Salary)	(7) ln(Salary)	(8) ln(Salary)
Panel A: NSCG 2015 Data								
Spread_2015	26.18	-71.33	-7.365	11.91	0.0010	-0.0003	0.0006	0.0008
	(93.45)	(73.99)	(71.99)	(73.35)	(0.0012)	(0.0009)	(0.0009)	(0.0009)
MCI_B_2015		12,821***				0.169***		
		(1,870)				(0.0191)		
MCI_W_2015			11,472***				0.155***	
			(1,928)				(0.0213)	
MCI_C_2015				10,739***				0.143***
				(2,282)				(0.0256)
Constant	65,798***	69,375***	67,028***	66,321***	11.04***	11.09***	11.06***	11.05***
	(4,331)	(3,314)	(3,373)	(3,398)	(0.0528)	(0.0386)	(0.0394)	(0.0404)
Observations	137	137	137	137	137	137	137	137
R-squared	0.001	0.522	0.426	0.374	0.007	0.517	0.441	0.382
Panel B: NSCG 2010 Data								
Spread_2010	67.02	-22.55	-0.624	22.895	0.00149	0.000179	0.000518	0.000848
	(87.01)	(78.75)	(72.49)	(75.77)	(0.00117)	(0.00103)	(0.000957)	(0.00100)
MCI_B_2010		7,801***				0.114***		
		(1,241)				(0.0165)		
MCI_W_2010			7,812***				0.112***	
			(1,434)				(0.0181)	
MCI_C_2010				6,775***				0.098***
				(1,250)				(0.0192)
Constant	65,260***	68,037***	67,357***	66,628***	11.05***	11.09***	11.08***	11.07***
	(3,431)	(3,164)	(3,001)	(3,110)	(0.0471)	(0.0420)	(0.0404)	(0.0420)
Observations	135	135	135	135	135	135	135	135
R-squared	0.005	0.238	0.243	0.186	0.013	0.254	0.250	0.205
Panel C: NSCG 2003 Data								
Spread_2003	115.9	115.6	111.6	128.2*	0.00172	0.00171*	0.00166*	0.00190*
	(80.00)	(71.27)	(70.87)	(74.82)	(0.00105)	(0.000923)	(0.000920)	(0.000971)
MCI_B_2003		8,489***				0.121***		
		(1,228)				(0.0161)		
MCI_W_2003			7,483***				0.108***	
			(1,532)				(0.0197)	
MCI_C_2003				6,265***				0.091***
				(1,663)				(0.022)
Constant	71,158***	71,170***	71,288***	70,790***	11.14***	11.14***	11.14***	11.13***
	(3,047)	(2,878)	(2,819)	(2,930)	(0.0395)	(0.0367)	(0.0357)	(0.0373)
Observations	140	140	140	140	140	140	140	140
R-squared	0.011	0.239	0.188	0.135	0.014	0.283	0.227	0.166

Note: Panel A reports results for the NSCG 2015 data; Panel B for the NSCG 2010 data; and Panel C for the NSCG 2003 data. The MCI measures are computed using 250 iterations. Robust standard errors are shown in parentheses. Significance is as follows: one-percent=***, five-percent=**, and ten-percent=*.

Table 3: Major Level Employment Rate Regressions

	(1) EmpRate	(2) EmpRate	(3) EmpRate	(4) EmpRate
Panel A: NSCG 2015 Data				
Spread_2015	0.0130	-0.0005	0.0072	0.0105
	(0.0218)	(0.0207)	(0.0194)	(0.0196)
MCI_B_2015		1.777***		
		(0.517)		
MCI_W_2015			1.970***	
			(0.373)	
MCI_C_2015				1.860***
				(0.404)
Constant	90.54***	91.03***	90.75***	90.62***
	(1.069)	(1.023)	(0.987)	(0.996)
Observations	137	137	137	137
R-squared	0.003	0.148	0.185	0.165
Panel B: NSCG 2010 Data				
Spread_2010	-0.0207	-0.0447	-0.0383	-0.0328
	(0.0282)	(0.0276)	(0.0254)	(0.0256)
MCI_B_2010		2.088***		
		(0.578)		
MCI_W_2010			2.027***	
			(0.411)	
MCI_C_2010				1.871***
				(0.396)
Constant	89.00***	89.74***	89.54***	89.37***
	(1.187)	(1.169)	(1.130)	(1.129)
Observations	135	135	135	135
R-squared	0.004	0.148	0.142	0.123
Panel C: NSCG 2003 Data				
Spread_2003	0.0270	0.0268	0.0256	0.0329
	(0.0303)	(0.0234)	(0.0255)	(0.0256)
MCI_B_2003		3.092***		
		(0.425)		
MCI_W_2003			2.476***	
			(0.404)	
MCI_C_2003				2.386***
				(0.422)
Constant	88.33***	88.34***	88.38***	88.19***
	(1.248)	(0.997)	(1.081)	(1.099)
Observations	140	140	140	140
R-squared	0.006	0.320	0.208	0.193

Note: Panel A reports results for the NSCG 2015 data; Panel B for the NSCG 2010 data; and Panel C for the NSCG 2003 data. MCI measures are computed using 250 iterations. Robust standard errors are shown in parentheses. Significance is as follows: one-percent=***, five-percent=**, and ten-percent=*.

Table 4: Pairwise Correlations Between the MCI and Major Specific Characteristics

Variable Description	MCI_B_2015
Standardized Test Scores	
SAT Verbal	0.36
SAT Mathematics	0.63
SAT Writing	0.29
Student Report - Developed Knowledge and Skills	
Writing clearly and effectively	-0.56
Speaking clearly and effectively	-0.59
Thinking critically and analytically	0.15
Analyzing quantitative problems	0.57
Using computing and information technology	0.52
Working effectively with others	-0.13
Learning effectively on your own	-0.14
Acquiring job or work-related knowledge and skills	0.05
Applying theories or concepts to practical problems or in new situations	0.46
Student Report - Time Spent	
Hours Spent Preparing for class	0.44
Amount of problem sets that take more than an hour to complete	0.63
Amount of problem sets that take less than an hour to complete	-0.17
Number of written papers or reports: 20 pages or more	0.14
Number of written papers or reports: between 5 and 19 pages	-0.24
Number of written papers or reports: fewer than 5 pages	-0.44

Note: Based on variation across 78 majors in the NSSE dataset for the years 2010-2011 that are mapped to 70 majors in the 2015 NSCG dataset.

Appendix A Building Block Model and Flow Network

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

Using the same example in Section 2 (Figure 1), there are four majors and four occupations. Now suppose the matching of students between majors and occupations are based on four latent skills. On the left-hand-side of Figure A.1, a link between a major and a skill indicates that this major equips 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 Skill 1 and 2 to be theorists. Since Skill 1 is only acquired in the Mathematics major, only students with a Mathematics degree can become theorists. In the similar logic, English and Fine Art majors cannot prepare students to be Accountants since Skill 2 is not taught in those majors.

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).

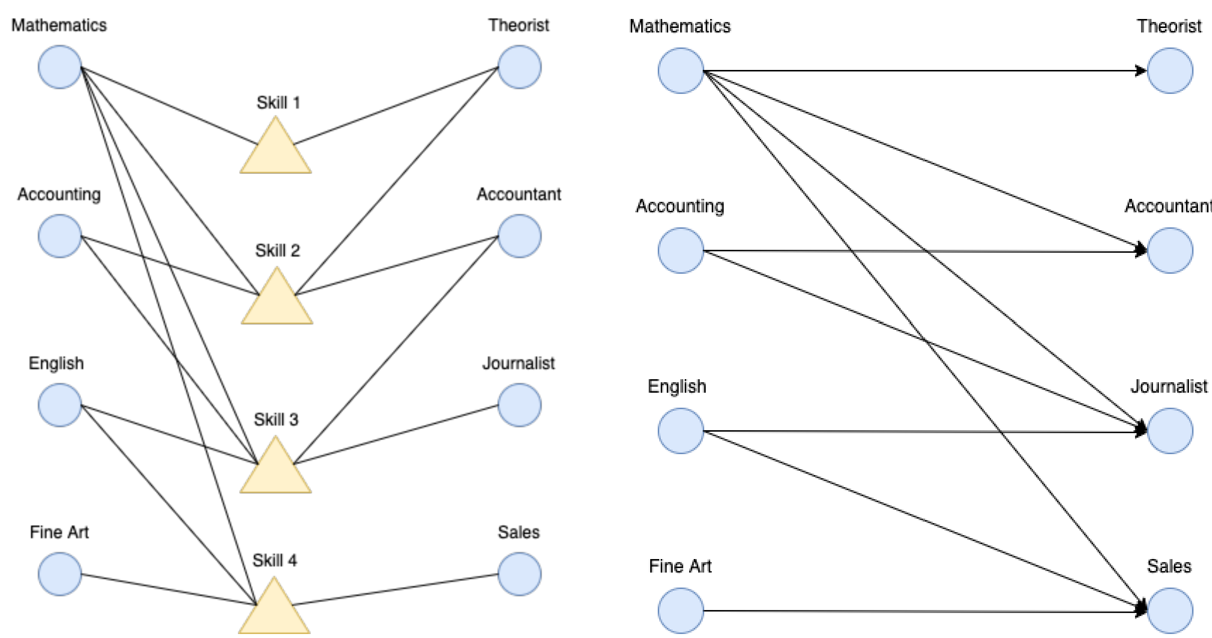

Figure A.1: Illustration of the "Building-Block" Model.

Appendix B Wage Regression Results Over Iterations

In applying the Method of Reflections to our wage regressions, we find that the explanatory power (measured by the R^2 value) of the MCI is increasing over iterations. For example, comparing column (2) and (6) in Table B.1, R^2 increases from 0.321 to 0.519 between the 2nd and 10th iteration, using the binary transition matrix and 2015 NSCG data. This is surprisingly large considering the Method of Reflections is merely a simple manipulation of the same data.

Table B.1: Wage Regression Results Over Iterations

	(1) Salary	(2) Salary	(3) Salary	(4) Salary	(5) Salary	(6) Salary
Spread_2015	26.18					
	(93.45)					
MCI_B_2015_iter2		9,954***				
		(1,665)				
MCI_B_2015_iter4			12,303***			
			(1,650)			
MCI_B_2015_iter6				12,631***		
				(1,692)		
MCI_B_2015_iter8					12,666***	
					(1,734)	
MCI_B_2015_iter10						12,658***
						(1,754)
Constant	65,798***	66,758***	66,758***	66,758***	66,758***	66,758***
	(4,331)	(1,242)	(1,076)	(1,048)	(1,045)	(1,046)
Observations	137	137	137	137	137	137
R-squared	0.001	0.321	0.490	0.516	0.519	0.519

Note: Binary MCI using NSCG 2015 Data. Note, Spread_2015 is the MCI value at iteration 0 (MCI_B_2015_iter0). Robust standard errors are shown in parentheses. Significance is as follows: one-percent=***, five-percent=**, and ten-percent=*.

Appendix C Major Complexity Index Rankings Over Time

Another interesting analysis based on our method is how the MCI ranking changes over time. Figure C.1 presents the dynamics of the major ranking using the Binary MCI over the 2003 to 2015 period. It reveals that there is considerable consistency in the major rankings where the year-on-year consistency of the ranking is qualitatively seen via the consistency of color-coding within Figure C.1 (i.e. red tends to stay on top, followed by orange, green, blue, and purple). Nevertheless, some majors have experienced substantial shifts across this time period. For instance, Actuarial Science was ranked 96th in 2003, but rose to 37th in 2010, and further to 24th in 2015. Interestingly, some majors change their rankings non-monotonically. It is plausible that these dynamics may reflect structural changes in the labor market. See Tables C.1-3 for more details of the MCI ranking changes over time (at CIP6 level).[17]

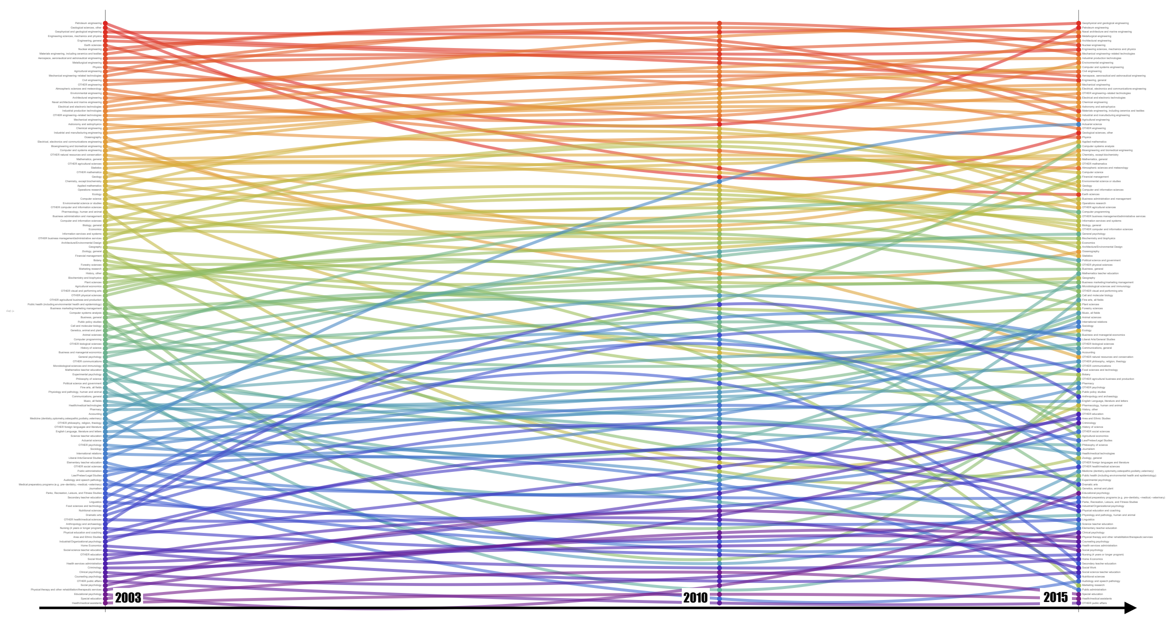

Figure C.1: Major Complexity Index (using the Binary MCI) Ranking Over Time (Years: 2003; 2010; 2015). Majors that are higher-up in the plot have higher MCI.

[17] For robustness check of major aggregation, Tables C.4-5 and Table C.6 report the major rankings using CIP4 and CIP2, respectively, with the Binary MCI and 2015 NSCG data.

Table C.1: Major Ranking at CIP6 Level

Major Description	(1) STEM	(2) Macro Field	(3) SATV	(4) SATM	(5) SATW	(6) 2003	(7) 2010	(8) 2015
Geophysical and geological engineering	1	Engineering				4	3	1
Petroleum engineering	1	Engineering				2	24	2
Mining and minerals engineering	1	Engineering				1		3
Naval architecture and marine engineering	1	Engineering				20	4	4
Metallurgical engineering	1	Engineering				11	7	5
Architectural engineering	1	Engineering				19	14	6
Nuclear engineering	1	Engineering				8	1	7
Engineering sciences, mechanics and physics	1	Engineering				5	8	8
Mechanical engineering-related technologies	1	Engineering				14	2	9
Industrial production technologies	1	Engineering				22	11	10
Environmental engineering	1	Engineering				18	22	11
Computer and systems engineering	1	Engineering				31	16	12
Civil engineering	1	Engineering	551	583	558	15	13	13
Aerospace, aeronautical and astronautical engineering	1	Engineering	605	655	584	10	9	14
Engineering, general	1	Engineering	583	633	570	6	10	15
Mechanical engineering	1	Engineering	574	632	556	24	17	16
Electrical, electronics and communications engineering	1	Engineering	573	633	555	29	18	17
OTHER engineering-related technologies	1	Engineering				23	21	18
Electrical and electronic technologies	1	Engineering				21	12	19
Chemical engineering	1	Engineering	598	660	585	26	19	20
Astronomy and astrophysics	1	Science				25	27	21
Materials engineering, including ceramics and textiles	1	Engineering	574	631	540	9	5	22
Industrial and manufacturing engineering	1	Engineering	561	615	517	27	20	23
Agricultural engineering	1	Engineering				13	6	24
Actuarial science	1	Business				100	37	25
OTHER engineering	1	Engineering				16	15	26
Geological sciences, other	1	Science				3	36	27
Physics	1	Science	625	653	584	12	23	28
Applied mathematics	1	Science				40	42	29
Computer systems analysis	1	Science				71	64	30
Bioengineering and biomedical engineering	1	Engineering				30	47	31
Chemistry, except biochemistry	1	Science	574	611	572	39	33	32
Mathematics, general	1	Science	579	642	583	34	31	33
OTHER mathematics	1	Science				37	45	34
Atmospheric sciences and meteorology	1	Science	569	596	557	17	30	35
Computer science	1	Science				43	32	36
Financial management	0	Business	536	584	536	57	57	37
Environmental science or studies	1	Agriculture	575	566	569	44	52	38
Geology	1	Science				38	26	39
Computer and information sciences	1	Science	570	600	544	48	28	40
Earth sciences	1	Science	566	570	563	7	34	41
Business administration and management	0	Business	510	528	496	47	48	42
Operations research	1	Engineering				41	25	43
OTHER agricultural sciences	1	Agriculture				35	35	44
Computer programming	1	Science				77	44	45
OTHER business management/administrative services	0	Business	483	494	450	53	29	46
Information services and systems	1	Science				52	43	47
Biology, general	1	Science	557	569	549	50	41	48
OTHER computer and information sciences	1	Science				45	39	49
General psychology	0	Social Sciences				81	54	50

Table C.2: Major Ranking at CIP6 Level

Major Description	(1) STEM	(2) Macro Field	(3) SATV	(4) SATM	(5) SATW	(6) 2003	(7) 2010	(8) 2015
Biochemistry and biophysics	1	Science	584	612	594	63	51	51
Economics	0	Social Sciences	592	614	597	51	50	52
Architecture/Environmental Design	1	Art & Humanities	552	591	556	54	40	53
Oceanography	1	Science				28	58	54
Statistics	1	Science	555	628	555	36	38	55
Political science and government	0	Social Sciences	589	564	579	87	53	56
OTHER physical sciences	1	Science	516	541	519	67	46	57
Business, general	0	Business				72	59	58
Mathematics teacher education	0	Education				84	96	59
Geography	1	Social Sciences	566	568	570	55	60	60
Business marketing/marketing management	0	Business	523	535	524	70	61	61
Microbiological sciences and immunology	1	Science	606	614	593	83	71	62
OTHER visual and performing arts	0	Art & Humanities	523	514	530	66	49	63
Cell and molecular biology	1	Science				74	82	64
Fine arts, all fields	0	Art & Humanities	551	535	543	88	106	65
Plant sciences	1	Agriculture				64	56	66
Forestry sciences	1	Agriculture				59	66	67
Music, all fields	0	Art & Humanities	567	555	552	91	87	68
Animal sciences	1	Agriculture				76	84	69
International relations	0	Social Sciences	505	527	487	103	90	71
Sociology	0	Social Sciences	531	521	530	102	81	72
Ecology	1	Science				42	76	73
Business and managerial economics	0	Business				80	69	74
Liberal Arts/General Studies	0	Art & Humanities	606	556	610	104	68	75
OTHER biological sciences	1	Science	568	575	550	78	79	76
Communications, general	0	Art & Humanities	535	525	533	90	77	77
Accounting	0	Business	523	570	514	94	63	78
OTHER natural resources and conservation	1	Agriculture	551	548	514	32	55	79
OTHER philosophy, religion, theology	0	Art & Humanities	602	576	592	96	85	80
OTHER communications	0	Art & Humanities				82	93	81
Food sciences and technology	1	Agriculture				116	72	82
Botany	1	Science	680	715	690	58	75	83
OTHER agricultural business and production	0	Agriculture	493	501	471	68	124	84
Pharmacy	1	Health	543	582	532	93	86	85
OTHER psychology	1	Social Sciences	541	529	532	101	97	86
Public policy studies	0	Social Sciences				73	117	87
Anthropology and archaeology	0	Social Sciences	586	555	569	120	65	88
English Language, literature and letters	0	Art & Humanities	604	551	599	98	89	89
Pharmacology, human and animal	1	Science				46	102	90
History, other	0	Art & Humanities	580	546	554	62	80	91
OTHER education	0	Education	503	503	497	127	109	92
Area and Ethnic Studies	0	Social Sciences	603	571	572	123	103	93
Criminology	0	Social Sciences				130	114	94
History of science	0	Art & Humanities				79	67	95
OTHER social sciences	0	Social Sciences	518	500	517	106	108	96
Agricultural economics	0	Agriculture				65	62	97
Law/Prelaw/Legal Studies	0	Social Sciences	523	507	505	108	88	98
Philosophy of science	0	Art & Humanities	636	596	625	86	94	99
Journalism	0	Art & Humanities	554	525	559	111	74	100

Table C.3: Major Ranking at CIP6 Level

Major Description	(1) STEM	(2) Macro Field	(3) SATV	(4) SATM	(5) SATW	(6) 2003	(7) 2010	(8) 2015
Health/medical technologies	1	Health	509	516	507	92	91	101
Zoology, general	1	Science	586	589	554	56	104	102
OTHER foreign languages and literature	0	Art & Humanities	602	579	602	97	73	103
OTHER health/medical sciences	1	Health	501	511	493	119	112	104
Medicine (dentistry,optometry,osteopathic,podiatry,veterinary)	1	Health	490	494	482	95	125	105
Public health (including environmental health and epidemiology)	1	Health				69	100	106
Experimental psychology	1	Social Sciences				85	132	107
Dramatic arts	0	Art & Humanities	576	557	568	118	92	108
Genetics, animal and plant	1	Science				75	70	109
Educational psychology	0	Social Sciences				136	130	110
Medical preparatory programs (e.g. pre-dentistry,-medical,-veterinary)	1	Health	513	529	499	110	110	111
Parks, Recreation, Leisure, and Fitness Studies	0	Other	501	517	485	112	120	112
Industrial/Organizational psychology	0	Social Sciences				124	99	113
Physical education and coaching	0	Education	475	490	456	122	101	114
Physiology and pathology, human and animal	1	Science				89	111	115
Linguistics	0	Art & Humanities				115	116	116
Science teacher education	0	Education				99	105	117
Elementary teacher education	0	Education	499	505	488	105	122	118
Clinical psychology	0	Social Sciences				131	128	119
Physical therapy and other rehabilitation/therapeutic services	0	Health	508	521	503	135	118	120
Counseling psychology	0	Social Sciences				132	119	121
Health services administration	0	Health				129	121	122
Social psychology	1	Social Sciences				134	113	123
Nursing (4 years or longer program)	0	Health	505	510	495	121	123	124
Home Economics	0	Social Sciences				125	126	125
Secondary teacher education	0	Education	528	524	510	114	83	126
Social Work	0	Social Sciences	480	469	460	128	115	127
Social science teacher education	0	Education				126	129	128
Nutritional sciences	1	Health				117	95	129
Audiology and speech pathology	0	Health	509	503	509	109	134	130
Marketing research	0	Business				60	78	131
Public administration	0	Social Sciences	558	551	579	107	107	132
Special education	0	Education	503	500	484	138	127	133
Health/medical assistants	0	Health				140	133	134
Education administration	0	Education				113		135
OTHER public affairs	0	Social Sciences				133	135	136
Counselor education and guidance services	0	Education				139		137
Science, unclassified	1	Science				33		
Data processing	1	Science				49		
Computer teacher education	0	Education				61		
Pre-school/kindergarten/early childhood teacher education	0	Education				137	131	

Note: Ranking based on the Binary MCI, using 2003, 2010, and 2015 NSCG data, sorted by the 2015 Binary MCI, at CIP6 level. Note, SAT scores are only available for those 70 NSCG majors using 2015 data that are mapped to the 78 NSSE majors. For cases where multiple NSSE majors are mapped to the same NSCG major, average SAT scores are reported. For instance, Civil engineering (SATV:547, SATM:606, SATW:537) and Urban planning (SATV: 554, SATM:559, SATW:579) from the NSSE are both linked to Civil engineering in the NSCG 2015. In this case, the average SAT scores (SATV: 551, SATM: 583, SATW: 558) are used for Civil engineering in the NSCG 2015, as shown in Table C.1.

Table C.4: Major Ranking at CIP4 Level

Major_CIP4	MCI_B_2015_Ranking
Geological/geophysical engineering	1
Mining and mineral engineering	2
Petroleum engineering	3
Naval architecture and marine engineering	4
Architectural engineering	5
Metallurgical engineering	6
Nuclear engineering	7
Engineering physics	8
Industrial production technologies/technicians	9
Mechanical engineering related technologies/technicians	10
Environmental/environmental health engineering	11
Systems engineering	12
Civil engineering	13
Aerospace, aeronautical and astronautical engineering	14
Engineering, general	15
Quality control and safety technologies/technicians	16
Mechatronics, robotics, and automation engineering	17
Electrical, electronics and communications engineering	18
Electromechanical instrumentation and maintenance technologies/technicians	19
Engineering chemistry	20
Manufacturing engineering	21
Astronomy and astrophysics	22
Polymer/plastics engineering	23
Management sciences and quantitative methods	24
Agricultural/biological engineering and bioengineering	25
Surveying engineering	26
Physics	27
Applied mathematics	28
Computer systems analysis	29
Chemistry	30
Biological/biosystems engineering	31
Mathematics	32
Atmospheric sciences and meteorology	33
Computer science	34
Finance and financial management services	35
Geological and earth sciences/geosciences	36
Computer and information sciences, general	37
Natural resources conservation and research	38
Business administration, management and operations	39
Operations research	40
Computer programming	41
General sales, merchandising and related marketing operations	42
Information science/studies	43
Biomathematics and bioinformatics	44
Computer/information technology administration and management	45
Agricultural public services	46
Psychology, general	47
Environmental design	48
Economics	49
Statistics	50
Biochemistry, biophysics and molecular biology	51
Political science and government	52
Nuclear and industrial radiologic technologies/technicians	53
Business/commerce, general	54
Film/video and photographic arts	55
Specialized sales, merchandising and marketing operations	56
Fine and studio art	57
Geography and cartography	58
Music	59
Microbiological sciences and immunology	60
Fire protection	61

Table C.5: Major Ranking at CIP4 Level

Major_CIP4	MCI_B_2015_Ranking
Animal sciences	62
Cell/cellular biology and anatomical sciences	63
Sociology and anthropology	64
Applied horticulture and horticultural business services	65
Business/managerial economics	66
International relations and affairs	67
Forestry	68
Ecology, evolution, systematics, and population biology	69
Accounting and related services	70
Liberal arts and sciences, general studies and humanities	71
Public relations, advertising, and applied communication	72
Theological and ministerial studies	73
Communication, journalism, and related programs, other	74
Clinical, counseling and applied psychology	75
Natural resources management and policy	76
Food science and technology	77
Teacher education and professional development, specific subject areas	78
Public policy analysis	79
Botany/plant biology	80
Agricultural production operations	81
Rhetoric and composition/writing studies	82
Anthropology	83
History	84
Social and philosophical foundations of education	85
Criminology	86
Area studies	87
Pharmacy, pharmaceutical sciences, and administration	88
Social sciences, other	89
Pharmacology and toxicology	90
Agricultural business and management	91
Non-professional general legal studies (undergraduate)	92
Journalism	93
American sign language	94
Alternative and complementary medicine and medical systems	95
Drama/theatre arts and stagecraft	96
Allied health diagnostic, intervention, and treatment professions	97
Zoology/animal biology	98
Public health	99
Research and experimental psychology	100
Teacher education and professional development, specific levels and methods	101
Medical residency programs - general certificates	102
Parks, recreation and leisure facilities management	103
Linguistic, comparative, and related language studies and services	104
Genetics	105
Health/medical preparatory programs	106
Physiology, pathology and related sciences	107
Health aides/attendants/orderlies	108
Health and medical administrative services	109
Family and consumer sciences/human sciences business services	110
Registered nursing, nursing administration, nursing research and clinical nursing	111
Social work	112
Marketing	113
Communication disorders sciences and services	114
Nutrition sciences	115
Public administration	116
Special education and teaching	117
Educational administration and supervision	118
Allied health and medical assisting services	119
Community organization and advocacy	120
Student counseling and personnel services	121

Note: Ranking based on the Binary MCI, using 2015 NSCG data, at CIP4 level.

Table C.6: Major Ranking at CIP2 Level

Major_CIP2	MCI_B_2015_Ranking
Engineering	1
Physical sciences	2
Engineering technologies/technicians	3
Mathematics and statistics	4
Natural resources and conservation	5
Computer and information sciences and support services	6
Social sciences	7
Biological and biomedical sciences	8
Business, management, marketing, and related support services	9
Psychology	10
Agriculture, agriculture operations, and related sciences	11
Visual and performing arts	12
Architecture and related services	13
Communication, journalism, and related programs	14
Security and protective services	15
Health professions and related clinical sciences	16
Education	17
Liberal arts and sciences, general studies and humanities	18
Science technologies/technicians	19
Theology and religious vocations	20
Area, ethnic, cultural, and gender studies	21
English language and literature/letters	22
History	23
Legal professions and studies	24
Public administration and social service professions	25
Residency programs	26
Foreign languages, literatures, and linguistics	27
Parks, recreation, leisure, and fitness studies	28
Multi/interdisciplinary studies	29
Family and consumer sciences/human sciences	30

Note: Ranking based on the Binary MCI, using 2015 NSCG data, at CIP2 level.

Appendix D NSSE Data Descriptive and Robustness Checks

To verify the robustness of major-level regression results presented in Table 2 and 3, we conduct further analysis controlling various features of college majors, including pre-college student characteristics such as SAT scores. To this end, we combine 2015 NSCG data with data from the National Survey of Student Engagement (NSSE) for the years 2010-2011 and compute average SAT scores of students in each major as well as major characteristics surveyed from students, such as knowledge and skills developed through college education and hours spent on coursework. We use data from these years since our final NSSE sample of 2010 and 2011 data include 43% freshmen and 57% seniors who are most likely in the labor market by the time of 2015. Table D.1 summarizes the variables used for the 78 majors that we are able to map between the two datasets.[18]

Table D.2 reports regression results where major-level features are controlled for. Comparing column (1) and (2) in Panel A, we see that while adding controls for average SAT scores reduces the impact of the MCI on mean salary, it still remains statistically and economically significant. A one standard deviation increase in the Binary MCI raises salary by $6,505 in column (2), which implies that after controlling for students' academic qualifications (i.e. removing potential positive selection bias on preexisting abilities), majors with higher complexity scores still produce substantially higher average earnings. Additional major characteristic controls in column (5) further reduce the MCI estimate down to $4,571, although, it is important to note that, because development of advanced knowledge and skills is the central channel through which the MCI affects earning outcomes, controlling for these additional characteristics may be an over-control for our purpose. Turning to Panel B of Table D.2, it is interesting that controlling for additional major features results in larger estimates of return (in terms of employment) to major complexity. For instance, a one standard deviation increase in the Binary MCI raises the employment rate by 1.77 percentage points in column (2), and 1.86 percentage points in column (5). Importantly, even with a limited sample size, these results indicate the robustness of the MCI in explaining the wage

[18]Specifically, we are able to map 78 majors in the NSSE to 70 majors in the NSCG. There are cases where multiple NSSE majors are mapped to the same NSCG major. For instance, Civil engineering and Urban planning (NSSE) are both linked to Civil engineering (NSCG).

and employment rate differentials across college majors.

Table D.3 displays the multivariate analysis of the Binary MCI on major-specific characteristics. Some of our intuition from the pairwise correlation remain consistent in the linear regressions: Firstly, as shown in column (1), SAT math is statistically significant at 1 percent level and positively correlated with the MCI. However, conditional on the math score, SAT verbal and writing are both negatively correlated with the MCI and not statistically significant from zero. Secondly, as shown in column (2), else equal, high MCI majors tend to have students report further development of thinking critically and analytically, and using computing and information technology; but less likely to report writing clearly and effectively, nor acquiring job or work-related knowledge and skills. Note, conditionally, analyzing quantitative problems is now negatively correlated with the MCI and not statistically significant, partly due to a high correlation with using computing and information technology (the correlation is 0.727, and the F-statistic of both variables is 8.89). Lastly, as shown in column (3), on average, high MCI majors tend to have students who report spending longer hours completing problem sets, and working on longer written assignments, ceteris paribus.

Table D.1: NSSE Summary statistics

Variable Description	Mean	Std. Dev.
Standardized Test Scores		
SAT Verbal	549.245	43.238
SAT Mathematics	559.034	51.246
SAT Writing	538.587	47.226
Student Report - Developed Knowledge and Skills		
Writing clearly and effectively	3.086	0.166
Speaking clearly and effectively	2.951	0.16
Thinking critically and analytically	3.36	0.09
Analyzing quantitative problems	3.094	0.231
Using computing and information technology	3.135	0.177
Working effectively with others	3.139	0.128
Learning effectively on your own	3.043	0.065
Acquiring job or work-related knowledge and skills	2.994	0.171
Applying theories or concepts to practical problems or in new situations	3.223	0.11
Student Report - Time Spent		
Hours Spent Preparing for class	4.447	0.398
Amount of problem sets that take more than an hour to complete	2.713	0.296
Amount of problem sets that take less than an hour to complete	2.542	0.204
Number of written papers or reports of 20 pages or more	1.472	0.132
Number of written papers or reports between 5 and 19 pages	2.427	0.215
Number of written papers or reports of fewer than 5 pages	3.047	0.174
N	78	

Note: Based on variation across 78 majors in the NSSE dataset for the years 2010-2011 that are mapped to 70 majors in the 2015 NSCG dataset.

Table D.2: Salary and Employment Rate Regressions controlling for Major Characteristics

Panel A:	(1) Salary	(2) Salary	(3) Salary	(4) Salary	(5) Salary
MCI_B_2015	8,442***	6,505***	5,948***	6,993***	4,571***
	(1,051)	(1,374)	(2,004)	(1,400)	(1,520)
SAT Verbal		-134.1			-27.09
		(84.32)			(122.5)
SAT Mathematics		150.5***			236.4*
		(48.87)			(119.8)
SAT Writing		-63.25			-107.8
		(64.02)			(87.95)
Student Report - Developed Knowledge and Skills			Yes		Yes
Student Report - Time Spent				Yes	Yes
Constant	64,698***	88,242***	32,924	-2,067	-112,731
	(1,168)	(14,959)	(67,768)	(43,602)	(87,897)
Observations	78	78	78	78	78
R-squared	0.404	0.520	0.557	0.643	0.723
Panel B:	EmpRate	EmpRate	EmpRate	EmpRate	EmpRate
MCI_B_2015	1.494***	1.770***	2.085***	1.405**	1.857***
	(0.343)	(0.495)	(0.629)	(0.545)	(0.651)
SAT Verbal		0.0245			-0.0739
		(0.0334)			(0.0508)
SAT Mathematics		-0.0135			0.0601
		(0.0152)			(0.0432)
SAT Writing		-0.0155			-0.00488
		(0.0237)			(0.0265)
Student Report - Developed Knowledge and Skills			Yes		Yes
Student Report - Time Spent				Yes	Yes
Constant	91.13***	93.58***	67.65***	111.8***	82.94***
	(0.404)	(6.442)	(20.89)	(16.98)	(30.40)
Observations	78	78	78	78	78
R-squared	0.151	0.161	0.388	0.232	0.491

Note: Panel A and B present results obtained by regressing major level average salaries and employment rate, respectively, on the Binary MCI using 2015 NSCG data and the major specific characteristics shown in Table D.1. Robust standard errors are shown in parentheses. Significance is as follows: one-percent=***, five-percent=**, and ten-percent=*.

Table D.3: Regression Between the MCI and Major Specific Characteristics

VARIABLES	(1) MCI_B_2015	(2) MCI_B_2015	(3) MCI_B_2015	(4) MCI_B_2015	(5) MCI_B_2015	(6) MCI_B_2015
Standardized Test Scores						
SAT Verbal	-0.00441			-0.00226	0.000919	-0.00259
	(0.00881)			(0.00960)	(0.00964)	(0.00876)
SAT Mathematics	0.0214***			0.0185**	0.00998*	0.0147*
	(0.00249)			(0.00780)	(0.00529)	(0.00758)
SAT Writing	-0.00818			-0.00865	-0.00499	-0.00676
	(0.00835)			(0.00705)	(0.00719)	(0.00804)
Student Report - Developed Knowledge and Skills						
Writing clearly and effectively		-2.212*		0.962		1.307
		(1.217)		(1.338)		(2.053)
Speaking clearly and effectively		-2.122		-2.459		-2.664*
		(1.518)		(1.577)		(1.571)
Thinking critically and analytically		3.468**		0.0439		-0.549
		(1.629)		(1.732)		(2.071)
Analyzing quantitative problems		-0.524		-1.862**		-1.446
		(0.695)		(0.932)		(1.053)
Using computing and information technology		3.219***		3.336***		1.783
		(0.832)		(0.895)		(1.096)
Working effectively with others		1.574		2.574*		1.291
		(1.405)		(1.378)		(1.614)
Applying theories or concepts to practical problems or in new situations		-1.086		0.997		0.687
		(1.294)		(1.277)		(1.271)
Acquiring job or work-related knowledge and skills		-2.343**		-3.242***		-2.269*
		(1.107)		(0.983)		(1.160)
Student Report - Time Spent						
Hours spent preparing for class			-0.187		-0.303	-0.231
			(0.285)		(0.258)	(0.337)
Amount of problem sets that take more than an hour to complete			2.382***		1.712***	1.197*
			(0.444)		(0.574)	(0.701)
Amount of problem sets that take less than an hour to complete			-2.446***		-1.776**	-0.894
			(0.653)		(0.744)	(0.819)
Number of written papers or reports: 20 pages or more			2.798***		2.317**	2.363**
			(0.892)		(0.921)	(1.167)
Number of written papers or reports: between 5 and 19 pages			-2.353**		-1.663*	-1.521
			(0.982)		(0.952)	(1.676)
Number of written papers or reports: fewer than 5 pages			1.036		0.949	1.170
			(1.070)		(1.001)	(1.174)
Constant	-5.142***	-1.462	-0.975	-6.559*	-4.443	-1.398
	(1.231)	(3.938)	(3.089)	(3.679)	(3.719)	(5.051)
Observations	78	78	78	78	78	78
R-squared	0.514	0.619	0.593	0.685	0.655	0.724
Adjusted R-squared	0.494	0.574	0.559	0.633	0.609	0.646

Note: Based on variation across 78 majors in the NSSE dataset for the years 2010-2011 that are mapped to 70 majors in the 2015 NSCG dataset. Robust standard errors are shown in parentheses. Significance is as follows: one-percent=***, five-percent=**, and ten-percent=*.

Appendix E Robustness Checks: CIP2 and CIP4

Table E.1: Major Level Wage Regressions - CIP4 and CIP2

	(1) Salary	(2) Salary	(3) Salary	(4) Salary	(5) ln(Salary)	(6) ln(Salary)	(7) ln(Salary)	(8) ln(Salary)
Panel A: NSCG 2015 Data_CIP4								
Spread_2015	-49.60	-112.61	-57.11	-37.73	-0.000	-0.001	-0.000	0.000
	(103.92)	(79.16)	(78.71)	(80.28)	(0.001)	(0.001)	(0.001)	(0.001)
MCI_B_2015		12,966***				0.168***		
		(2,010)		.		(0.021)		
MCI_W_2015			11,518***				0.154***	
			(1,987)				(0.022)	
MCI_C_2015				10,704***				0.142***
				(2,355)				(0.026)
Constant	70,140***	72,595***	70,433***	69,678***	11.103***	11.135***	11.107***	11.097***
	(5,022)	(3,742)	(3,878)	(3,910)	(0.060)	(0.044)	(0.045)	(0.046)
Observations	121	121	121	121	121	121	121	121
R-squared	0.003	0.521	0.416	0.359	0.000	0.508	0.430	0.364
Panel B: NSCG 2015 Data_CIP2								
Spread_2015	191.14*	-281.50*	40.00	50.87	0.003*	-0.004*	0.001	0.001
	(109.28)	(150.48)	(92.76)	(93.41)	(0.002)	(0.002)	(0.001)	(0.001)
MCI_B_2015		11,470***				0.175***		
		(2,614)				(0.039)		
MCI_W_2015			7,136***				0.104***	
			(1,388)				(0.020)	
MCI_C_2015				6,855***				0.099***
				(1,302)				(0.019)
Constant	50,993***	79,036***	59,961***	59,316***	10.850***	11.278***	10.981***	10.971***
	(6,697)	(9,325)	(6,186)	(6,243)	(0.104)	(0.140)	(0.096)	(0.097)
Observations	30	30	30	30	30	30	30	30
R-squared	0.136	0.377	0.467	0.446	0.142	0.385	0.446	0.424

Note: Panel A reports results for the NSCG 2015 data at CIP4 level; Panel B for the NSCG 2015 data at CIP2 level. The MCI measures are computed using 250 iterations. Robust standard errors are shown in parentheses. Significance is as follows: one-percent=***, five-percent=**, and ten-percent=*. The estimates of MCI_C_2015 are similar when additional variables (i.e. race and parental education in addition to age and gender) are controlled for using CIP2: 6,660*** (1,277) and 0.097*** (0.018) compared to the ones shown in Panel B, column (4) and (8), respectively.

Table E.2: Major Level Employment Rate Regressions - CIP4 and CIP2

	(1) EmpRate	(2) EmpRate	(3) EmpRate	(4) EmpRate
Panel A: NSCG 2015 Data_CIP4				
Spread_2015	0.014	0.005	0.013	0.017
	(0.023)	(0.021)	(0.020)	(0.020)
MCI_B_2015		2.006***		
		(0.558)		
MCI_W_2015			2.088***	
			(0.395)	
MCI_C_2015				1.979***
				(0.428)
Constant	90.604***	90.983***	90.657***	90.518***
	(1.199)	(1.094)	(1.072)	(1.084)
Observations	121	121	121	121
R-squared	0.004	0.200	0.219	0.197
Panel B: NSCG 2015 Data_CIP2				
Spread_2015	0.074**	-0.051	0.039	0.041
	(0.029)	(0.051)	(0.026)	(0.026)
MCI_B_2015		3.023***		
		(0.900)		
MCI_W_2015			1.660***	
			(0.448)	
MCI_C_2015				1.634***
				(0.444)
Constant	85.935***	93.326***	88.021***	87.918***
	(1.964)	(3.190)	(1.814)	(1.814)
Observations	30	30	30	30
R-squared	0.231	0.421	0.434	0.431

Note: Panel A reports results for the NSCG 2015 data at CIP4 level; Panel B for the NSCG 2015 data at CIP2 level. MCI measures are computed using 250 iterations. Robust standard errors are shown in parentheses. Significance is as follows: one-percent=***, five-percent=**, and ten-percent=*. The estimate of MCI_C_2015 is similar when additional variables (i.e. race and parental education in addition to age and gender) are controlled for using CIP2: 1.657** (0.445) compared to the one shown in Panel B, column (4).

Appendix F Robustness Checks: Traditional Major Category

Table F.1: Salary and Employment Rate Regressions controlling for Traditional Major Category

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A:	Salary	Salary	Salary	Salary	Salary	Salary
MCI_B_2015	12,622***			13,275***	10,327***	10,628***
	(1,767)			(2,242)	(3,432)	(3,560)
STEM		13,269***		-2,248		-2,982
		(2,672)		(2,775)		(2,706)
Art & Humanities			-459.6		2,526	479.9
			(1,826)		(2,572)	(2,889)
Business			24,779***		22,868***	20,825***
			(3,171)		(3,039)	(3,637)
Education			-4,987***		4,941	2,912
			(1,810)		(4,696)	(4,502)
Engineering			32,949***		15,723***	15,884***
			(4,387)		(4,881)	(4,755)
Health			6,263*		14,924***	14,596***
			(3,483)		(4,690)	(4,659)
Other			3,600***		11,454***	9,364***
			(1,252)		(3,339)	(3,202)
Science			10,974***		7,148**	7,699***
			(2,363)		(3,022)	(2,755)
Social Sciences			2,348		7,460**	5,787*
			(2,529)		(3,305)	(3,047)
Constant	66,758***	59,397***	55,906***	68,005***	57,353***	59,715***
	(1,049)	(1,505)	(1,252)	(1,989)	(2,059)	(3,219)
Observations	137	137	137	137	137	137
R-squared	0.516	0.142	0.526	0.518	0.619	0.620
Panel B:	EmpRate	EmpRate	EmpRate	EmpRate	EmpRate	EmpRate
MCI_B_2015	1.776***			2.151***	1.299	1.419
	(0.497)			(0.646)	(1.025)	(1.071)
STEM		1.225		-1.289		-1.183
		(0.772)		(0.935)		(1.367)
Art & Humanities			-1.629		-1.253	-2.065
			(1.742)		(1.799)	(1.985)
Business			-0.123		-0.363	-1.174
			(1.691)		(1.621)	(1.974)
Education			-2.900		-1.651	-2.456
			(2.410)		(2.459)	(2.590)
Engineering			2.229		0.0609	0.125
			(1.549)		(2.035)	(2.014)
Health			-3.758**		-2.669	-2.798
			(1.771)		(2.048)	(1.962)
Other			-1.999		-1.010	-1.839
			(1.320)		(1.604)	(1.802)
Science			-1.919		-2.401	-2.182
			(1.619)		(1.710)	(1.670)
Social Sciences			-1.957		-1.314	-1.978
			(1.441)		(1.583)	(1.809)
Constant	91.01***	90.33***	92.11***	91.73***	92.29***	93.23***
	(0.365)	(0.519)	(1.320)	(0.593)	(1.385)	(1.786)
Observations	137	137	137	137	137	137
R-squared	0.148	0.018	0.161	0.161	0.183	0.187

Note: Panel A and B present results obtained by regressing major level average salaries and employment rate, respectively, on the Binary MCI using 2015 NSCG data. There are totally 9 macro fields of study where Agriculture is the omitted group. See Tables C.1-3 for detailed information on the macro fields. Robust standard errors are shown in parentheses. Significance is as follows: one-percent=***, five-percent=**, and ten-percent=*.