

# COLLEGE MAJOR CHOICE AND NEIGHBORHOOD EFFECTS IN A HISTORICALLY SEGREGATED SOCIETY: EVIDENCE FROM SOUTH AFRICA

**Biniam E. Bedasso**

Collaborative Africa Budget  
Reform Initiative  
Centurion, 0062 South Africa  
biniam.bedasso@cabri-sbo.org

## Abstract

This paper explores factors affecting the choice of investment in specific human capital in the presence of significant inter-group and spatial inequalities. I use four years of admissions application data at an elite university in South Africa in conjunction with quarterly labor force data to trace the link between aptitude-adjusted expected earnings, neighborhood effects, and the choice of college major. The paper relies on the availability of a rich set of academic and geographical information in the admissions database to make causal inference. The results show that expected earnings have a positive impact on major choice independently of high school background when the ex ante distribution of earnings captures the full range of between-major and within-major income differentials. White applicants are more responsive to differentials in expected earnings than black applicants. Neighborhood effects influence college major choice through near-peer role models and relative achievement at the high school level.

[https://doi.org/10.1162/edfp\\_a\\_00249](https://doi.org/10.1162/edfp_a_00249)

© 2019 Association for Education Finance and Policy

## 1. INTRODUCTION

In an era of increasing specializations and rising wage differentials, not all diplomas are created equal. There are a number of studies in the existing literature that show the link between the distribution of college majors and income inequality (Grogger and Eide 1995; Arcidiacono 2004). Gender and racial income gaps are also partly explained by the heterogeneity of human capital (Daymont and Andrisani 1984; Weinberger 1998). As one goes up the education ladder, differences in income will be as much about the type of training acquired as about the number of years spent in school. Hence, understanding inequality, especially in the middle part of the income distribution, requires having a good grasp of factors influencing the allocation of heterogeneous human capital. The impact of the allocation of specialized training is likely to be pronounced in developing countries where it has more power to shape the composition of elites and the nature of the middle class. The allocation of talent across various fields of specialization with differential impact on innovation and technological change will affect growth in the long run (Murphy, Shleifer, and Vishny 1991). The effect of the allocation of specialized human capital could be far-reaching enough to influence institutional quality to the extent that it shapes the preferences of elites (Bedasso 2015).

This paper examines the determinants of college major choice in South Africa in the context of significant inter-group and spatial inequalities that have continued to characterize South African society.<sup>1</sup> By investigating factors influencing college major choice, I seek to shed light on the role of heterogeneous human capital in the reproduction of inequality in South Africa. Specifically, the paper attempts to answer two questions: (1) How do various groups respond to differentials in major-specific expected earnings against the backdrop of sizeable inequality in economic and sociopolitical endowments? and (2) How do peer effects and relative achievement influence college major choice at the levels of specific neighborhoods and high schools? I exploit the extensive information contained in the admissions database of the University of Cape Town (UCT) between 2010 and 2013, jointly with the Quarterly Labor Force Survey conducted at a national level. The novelty of the dataset lies in the amount of information it provides about the high school education and area of residence of college applicants. Moreover, the status of UCT as the best ranked institution of higher education on the African continent allows me to put the analysis in the context of elite formation in a society that has been undergoing social and political transformation.<sup>2</sup>

In standard theoretical frameworks, college major choice is analyzed as part of a lifecycle model of stochastic career choice (Altonji, Blom, and Meghir 2012). This approach helps establish a link between educational choices at earlier stages in life and the choice of college major. Therefore, to the extent that pre-college educational opportunities are determined by space, it is possible to draw a connection between spatial inequality and major choice. There are a number of channels through which spatial

- 
1. The educational inequalities of the Apartheid era continue to persist in South Africa, as manifested in disparities in schooling outcomes between historically black and historically white schools (Van der Berg 2007). Regional inequality is rife in the schooling system in South Africa. As of 2013, less than 44 percent of public schools in one of the poorest provinces offer math in grades 10 through 12. The corresponding figure for the best served province is 91 percent.
  2. Two of the most widely cited university rankings—the Times Higher Education Ranking and the QS World University Ranking—consistently rank the University of Cape Town as the best university in Africa.

inequality may influence individual decisions in college. This includes the quality of schools in a given geographical area, the influence of near-peer role models, and the effect of relative achievement in different schools. Essentially, individuals are constrained by all or some of these background factors as they optimize expected lifetime earnings from each major.

I estimate a random utility model of the determinants of choice between five faculties (or departments) at UCT with nesting structure elicited from the data. I exploit the availability of two sets of national test scores with varying emphasis on measuring overall ability and college adaptability to identify the effect of aptitude-adjusted expected earnings while controlling for academic ability directly. I use two specifications of expected earnings based on alternative assumptions regarding the information set applicants face about future income. The results show that aptitude-adjusted expected earnings exert a positive impact on major choice. However, the effect of expected earnings is absorbed by the antecedent effect of high school curriculum choice when the *ex ante* distribution of earnings is assumed to be egalitarian. On the contrary, when the *ex ante* distribution of earnings accounts for merit-based differentials in income, expected earnings continue to hold a positive effect on major choice independently of high school curriculum. As far as racial disparities are concerned, white applicants are shown to be, on average, 1.8 times more responsive to major-specific earnings differentials than black applicants.

With regard to neighborhood effects, I exploit the variation in past admission trends to UCT across 3,773 postcodes represented in the database to capture the influence of near-peer role models. Correcting for possible clustering of unobserved preferences along postcodes, a one standard deviation increase in the ratio of near-peers who were admitted to a certain faculty during the last three years is shown to increase by around 9 percent the probability of choosing the same faculty. I also endeavor to explore whether there is a bright side to spatial inequality. Based on the assumption that applicants who belong to the high end of the grade distribution in a less competitive high school have better self-esteem than academically similar students who have gone to more competitive high schools, I test whether being a “big frog in a small pond” induces applicants to choose high return majors. The results show that top ranking applicants from a bottom quartile high school are consistently more likely to choose health and other sciences over humanities than applicants with higher average grades who nevertheless belong to a lower rung of the grade distribution in a top quartile high school.

This paper belongs in a long vein of literature on career choice under uncertainty (see for example, Altonji 1993; Keane and Wolpin 1999; Arcidiacono 2004; Montmarquette, Cannings, and Mahseredjian 2002; Wiswall and Zafar 2015). For that matter, much of the theoretical framework I formulate in the next section is based on these papers. Most studies have, either implicitly or explicitly, dealt with the impact of major choice on income inequality. However, there has been little focus on the implications of overall inequality, let alone group and spatial inequality, for the pattern of major choice. To be sure, there have been notable theoretical contributions linking segregation and human capital investment to overall inequality (see, for example, Bénabou 1996). The economics and sociology literatures are rich in empirical evidence on the impact of geographical segregation on educational outcomes (Braddock 1980; Card and Rothstein 2007; Goldsmith 2009). But there is still little evidence on the effect of neighborhood

characteristics on educational outcomes in the context of heterogeneous human capital. This paper contributes to the literature in three ways. First, it takes advantage of extensive geographical variation in the data to analyze the determinants of college major choice in perspective with spatial inequality. Second, it enriches the empirical measurement of expected earnings by using alternative assumptions regarding the ex ante distribution of future income. Third, it exploits big administrative data to estimate the determinants of major choice in a developing country, which would otherwise have been impossible due to the absence relevant survey data.<sup>3</sup>

The remainder of the paper is organized as follows. Section 2 outlines the theoretical framework and the empirical strategy I apply to answer the questions raised in this paper. Section 3 provides an overview of the data and summary statistics. Section 4 is devoted to presentation of results. Section 5 concludes.

## 2. THEORETICAL FRAMEWORK AND EMPIRICAL STRATEGY

The choice of college major is part of a dynamic lifecycle decision-making process. In most cases, the sequence of choices starts as early as high school, where students and their parents may have to choose the type of curriculum to be pursued. This will then be followed by a series of choices on whether to go to college or join the labor market, which college to attend, and what major to choose. The last stage of decision-making often consists of the choice of occupation conditional on education. Because I am dealing with the choice of college major at the point of admission, I will start with specifying the value function of individual  $i$  who has already chosen to go to college and picked out major  $j$  from a set of available alternatives  $j = 1, 2, \dots, J$ ,

$$v_{ij} = E_1(u_{ij} | a, \theta) + \beta \int E_1(v_{2ij} | a) da, \quad (1)$$

where  $u_{ij}$  is the flow utility in period 1 (i.e., college) conditional on ability,  $a$ , and preference,  $\theta$ , both of which might not be directly observable. The second term in equation 1 comprises the terminal value of lifetime earnings expected in period 2 (i.e., work) conditional on ability.  $\beta$  is the discount rate. The utility received by the student while attending college in major  $j$  is determined by the psychic and pecuniary costs of major  $j$ , as well as individual preference for that particular major. This can be written as

$$u_{ij} = \alpha_1 C_{ij} + \alpha_2 X_i + \varepsilon_{ij}, \quad (2)$$

where  $X_i$  is a vector of individual, school, and neighborhood characteristics that may influence the inclination of student  $i$  towards major  $j$ .  $\varepsilon_{ij}$  is the unobserved preference of the individual.  $C_{ij}$  is a combination of psychic and pecuniary costs specific to the individual. The individualized cost function is given by

$$C_{ij} = \gamma_1 A_{ij}(h) + \gamma_2 Z_{ij}, \quad (3)$$

where  $A_{ij}(h)$  is the observed academic ability of the individual for major  $j$  that is assumed to affect the psychic cost of major  $j$  through the intensity of effort required to

3. Most of the existing studies of college major choice are based on the National Longitudinal Surveys of Youth in the United States.

complete college. At the point of college admission, observed ability is a function of high school preparation,  $h$ , and, therefore, conditional on choices made earlier in life.  $Z_{ij}$  is the pecuniary cost of major  $j$ , which is assumed to be specific to individual  $i$  because the relative cost of college varies across socioeconomic backgrounds. This is because credit markets are imperfect. Substituting equation 3 in equation 2, the flow utility of attending college in major  $j$  can be parameterized as follows,

$$u_{1ij} = \alpha_1 \gamma_1 A_{ij}(h) + \alpha_1 \gamma_2 Z_{ij} + \alpha_2 X_i + \varepsilon_{1ij}. \quad (4)$$

The measure of expected lifetime earnings—that is, the second term of the value function in equation 1—depends on the type of information applicants have, *ex ante*, about earnings differentials across and within college majors. Accordingly, I formulate two discrete form specifications of expected earnings based on alternative assumptions regarding the information set applicants face. First, assuming that applicants have generic information about the returns to major  $j$ , the expected earnings of individual  $i$  in major  $j$  can be written as

$$E_1(v_{2ij}) = p_{ij}(a)e_j^g + (1 - p_{ij}(a))e_j^d, \quad (5)$$

where  $p_{ij}(a)$  is the probability of successfully completing college, which is assumed to be a function of ability ( $a$ ), and  $e_j^g$  and  $e_j^d$  represent, respectively, median income after graduating in major  $j$  and median income after dropping out of college. Because applicants are assumed to have generic information about earnings differentials, ability affects expected earnings, *ex ante*, only through the probability of successfully completing college. In a way, this specification can be considered to be based on egalitarian distribution of expected earnings.

Second, assuming that applicants have full information, *ex ante*, about where they are likely to fall in the earnings distribution based on their ability, the expected earnings of individual  $i$  with ability  $a$  in major  $j$  is given as

$$E_1(v_{2ij}) = e_{j,a}^g. \quad (6)$$

In contrast to the egalitarian measure of expected earnings in equation 5, this specification can be considered to be predicated on meritocratic distribution of expected earnings.

I can now combine equations 4 and 5 or 6 to arrive at a reduced form equation that can be estimated empirically. This means the indirect utility for individual  $i$  from majoring in  $j$  can be written as a linear function of observed academic ability, individual and neighborhood characteristics, and expected earnings:

$$v_{1ij} = \alpha_1 \gamma_1 A_i(h) + \alpha_1 \gamma_2 Z_{ij} + \alpha_2 X_i + \mu_\beta \overline{E_1(v_{2ij})} + \varepsilon_{1ij}, \quad (7)$$

where  $\overline{E_1(v_{2ij})} = p_{ij}(a)(e_{ij}^g - e_{ij}^d) + p_{ij}e_{ij}^d$  or  $\overline{E_1(v_{2ij})} = e_{j,a}^g$ .

Generally, the choices made in high school, which are expected to influence major choice through  $A_{ij}$ , are constrained by geographical space. In countries such as South Africa that are characterized by significant spatial inequality, neighborhood characteristics affect major choice both directly through shaping preferences and indirectly through  $A_{ij}(h)$ . Hence, provided there are appropriate constructs to measure regional

distribution of resources and neighborhood characteristics, the model specified in equation 7 can be estimated to explain the choice of college major in South Africa.

The parameters in equation 7 can be estimated using standard techniques as long as certain assumptions are made regarding the distribution of unobservables. First, I assume that  $\varepsilon_{ij}$  is independently distributed across individuals once relevant neighborhood characteristics are controlled for. Second, I assume  $\varepsilon_{ij}$  is correlated across majors sharing certain characteristics. For example, the unobserved preference of a student for a physics degree is likely to be correlated with the unobserved preference for a chemistry degree. The same might not hold between physics and history. Formally, suppose there are  $N$  groups of academic streams across which the  $J$  majors are distributed. This means a student will have to choose an academic stream  $n = 1, \dots, N$  before deciding on which specific major to pursue. This implies that the final choice set can be written as follows:

$$j \in \{(j_{11}, \dots, j_{J,1}), (j_{12}, \dots, j_{J,2}), \dots, (j_{1N}, \dots, j_{J,N})\}.$$

This type of structure leads to nested logit probabilities with  $J$  alternatives and  $N$  nests. Suppose the full set of explanatory variables can be split into explanatory variables of nest choice,  $W_n$ , and explanatory variables of the choice of specific alternatives,  $S_{nj}$ . Note that I have dropped the individual index  $i$  and the period index 1 for simplicity. This means equation 7 can be rewritten as

$$v_j = W'_n \kappa_n + S'_{nj} \lambda_j + \varepsilon_{nj}. \tag{8}$$

Provided that  $\varepsilon_{nj}$  is drawn from a generalized extreme value distribution, the probability of choosing the  $j^{th}$  alternative belonging to the  $n^{th}$  nest is given by

$$\begin{aligned} \Pr(j) &= \Pr(j | n) \Pr(n) \\ &= \frac{\exp(S'_{nj} \lambda_j / \tau_n)}{\sum_{j \in J_n} \exp(S'_{nj} \lambda_j / \tau_n)} \left( \frac{\exp[W'_n \kappa_n + \tau_n IV_n]}{\sum_{n \in N} \exp[W'_n \kappa_n + \tau_n IV_n]} \right), \end{aligned} \tag{9}$$

where  $IV_n = \ln(\sum_{j=1}^{J_n} \exp(S'_{nj} \lambda_j))$  and  $\tau_n$  measures the correlation of  $\varepsilon_{nj}$  within a given nest.

The coefficients in equation 9, including the dissimilarity coefficient  $\tau_n$ , can be estimated using the full information maximum likelihood method. I will return to the issue of identification once I have presented the data in the next section.

### 3. DATA

I use the admissions application database of the UCT to estimate the model of major choice specified in the previous section. The data are available for the entire population of applicants, which ranges between 13,897 and 16,077 applicants per year, for the four years from 2010 to 2013. In addition to basic demographic information and academic record, as of 2013, UCT has started collecting data on family background and high school characteristics from all applicants. Therefore, I use the population of applicants in 2013 to estimate the model. Moreover, information from previous years is utilized to identify neighborhood trends based on patterns of admission of applicants

from a given neighborhood over the four years. The application database is rich in geographical information as applicants could be traced to the level of residential postcode. Moreover, the names of the high schools of the applicants are given in the dataset. This means I can exploit spatial variations across 3,773 postcodes and 1,577 high schools.

Much of the analysis is based on the entire population of applicants, regardless of admission status. This is done in the interest of minimizing the selection bias that may arise due to the university admission process, as well as the decision of applicants whether or not to accept offers. Prospective students apply for majors in one or two of UCT's faculties as their first and second choices. Applications to individual faculties are evaluated independently and admission is offered or denied by the respective faculty. Therefore, the first choices of prospective students at the point of application are supposed to reflect the revealed preferences of individuals before supply constraints are imposed. Second choices, by contrast, are less informative because applicants may act strategically by listing a less popular major as a second choice in order to minimize potential loss.<sup>4</sup>

Estimation of the model specified in the previous section requires data on expected earnings. The combination of earnings and college major data are available for the third and fourth quarters of the Quarterly Labor Force Survey in 2012. Ideally, the expected lifetime earnings of applicants in each field of study are estimated using a sample of current workers of various age groups who share similar characteristics as the applicant. However, this approach can be problematic in the case of South Africa because the earnings trajectories of mature workers in 2012 are likely to have been shaped by a distorted system during apartheid. This dynamic is unlikely to hold in the postapartheid period. Therefore, I have restricted the age limit of the sample of workers that are used to calculate expected earnings to thirty years.<sup>5</sup> Another challenge in this case is that there are not enough observations of college educated, under-thirty-years-olds in the Quarterly Labor Force Survey data to estimate earnings on a full range of personal and group characteristics. Therefore, I have chosen to use the earnings of the under-thirty-year-olds sample for each major to approximate expected earnings of college applicants. The fact that the earnings of current workers are based on realized choices of college major might introduce selection bias in the sense that current workers are already sorted by aptitude. This means the expected earnings of high-scoring applicants in less selective majors could be biased downwards. The bias is upward in case of low-scoring applicants in more selective majors. However, I expect the effect of such biases to be minimal since the earnings differential among majors is more likely to emanate from systematic differences in the occupations in which they lead than from individual ability. I will return to the calculation of individualized aptitude-adjusted expected earnings later in this section.

Table 1 presents a summary of variables that are used to estimate the model of major choice. The categories of choice are organized in terms of official faculties at UCT. I have combined the faculties of humanities and law to simplify the choice set into five faculties: three in science and technology (health sciences, engineering, science) and two in

4. I would like to thank one of the anonymous referees for pointing out this scenario.

5. This choice makes more practical sense particularly if one assumes that college students are able to observe more closely the incomes of those who are in the same generation as they are than that of older people.

**Table 1.** Summary Statistics of College Applications and Labor Market Data

	Commerce	Engineering	Health Sciences	Science	Humanities and Law	Total
Number of applicants	3,359	2,600	3,738	1,263	3,257	14,217
Ratio of admitted	0.22	0.12	0.08	0.21	0.20	0.15
Academic record						
NSC Math	67.8 (17.2)	69.0 (16.4)	65.9 (17.5)	64.1 (17.1)	59.5 (18.1)	65.4 (17.7)
NSC English	70.5 (9.9)	68.5 (9.8)	71.2 (10.2)	68.8 (9.8)	66.6 (10.5)	69.3 (10.3)
NSC Life orientation	79.6 (9.3)	78.3 (9.3)	80.8 (8.9)	77.9 (9.2)	75.1 (10.0)	78.4 (9.6)
NBT Math	46.9 (17.3)	48.2 (17.2)	43.6 (16.8)	43.5 (16.4)	36.0 (12.7)	44.7 (17.0)
NBT Academic literacy	63.5 (12.7)	60.2 (14.3)	60.2 (13.9)	61.1 (14.1)	60.9 (13.7)	61.2 (13.8)
Number of high school science courses	1.5 (1.1)	2.4 (0.81)	2.3 (0.71)	2.3 (0.78)	1.1 (0.90)	1.8 (1.0)
Socioeconomic background						
Ratio of public school	0.27	0.33	0.31	0.31	0.27	0.29
Ratio of private school	0.73	0.67	0.69	0.69	0.73	0.71
Ratio of first-generation college	0.42	0.47	0.48	0.51	0.47	0.47
Demographic variables						
Male	0.48	0.70	0.31	0.47	0.34	0.44
Female	0.52	0.30	0.69	0.53	0.66	0.56
Black	0.50	0.55	0.53	0.51	0.48	0.51
White	0.24	0.22	0.17	0.26	0.25	0.24
Colored and Asian	0.25	0.21	0.28	0.22	0.25	0.25
Labor market variables						
Undergraduate median earnings: Under 30 years old (in rand)	16,000	20,000	18,000	16,000	15,000	15,000
Technical school median earnings: Under 30 years old (in rand)	10,200	11,000	10,000	10,000	9,000	10,500
Expected earnings 1: Egalitarian distribution (in rand)	11,939 (738)	13,195 (1,978)	12,582 (1,364)	10,545 (1,375)	9,449 (555)	11,487 (1,879)
Expected earnings 2: Meritocratic distribution (in rand)	19,399 (13,831)	21,760 (13,311)	20,023 (12,015)	17,912 (12,269)	16,662 (11,746)	19,079 (12,791)

Notes: Standard deviation is given in parentheses in case of mean. NBT = National Benchmark Test; NSC = National Senior Certificate.

social sciences (commerce, humanities and law). The admission rates show that health sciences and engineering are by far the most selective faculties. Academic preparation is measured by six variables displayed in the second panel of table 1. First, scores in the three compulsory subjects of the National Senior Certificate (NSC) examination—English, mathematics, and life orientation—are applied to measure basic academic ability. Second, scores in the National Benchmark Test (NBT), which is used as an additional requirement for admission into major universities in South Africa, is utilized to measure the probability of success in college and in a professional environment. The NBT is specifically designed to gauge the adaptability of students to college curriculum. The test is given in three modules: academic literacy, quantitative literacy, and mathematics. Third, the number of science courses a student has taken in high school is used as a proxy for the level of preparation in science and technology. The high



standard deviation, relative to the mean of the number of science courses, indicates that high school graduates vary significantly in terms of their preparation in science.

Table 1 also presents variables that are used to measure socioeconomic background. A majority of the applicants attended private high schools, signifying the middle-class status of most applicants. In terms of the educational background of families, over 47 percent of applicants would be the first ones in three generations to have earned a college degree. Demographic characteristics of applicants show that women constitute the majority of applicants. Black applicants form 51 percent of the applicant pool. White applicants constitute 24 percent and Colored applicants (people of mixed ancestry, which has become a distinctive group through the lines drawn by the Apartheid regime) and Indians make up 25 percent.

The bottom panel in table 1 presents summary of expected earnings for each faculty. The first row contains the median monthly earnings of workers under thirty years of age with a three- or four-year college degree in each field. The second row presents the median income of workers in the same age group who have completed only a technical school diploma in one of the five fields. The incomes of technical school graduates are utilized to approximate the expected earnings of applicants in case they drop out of college.

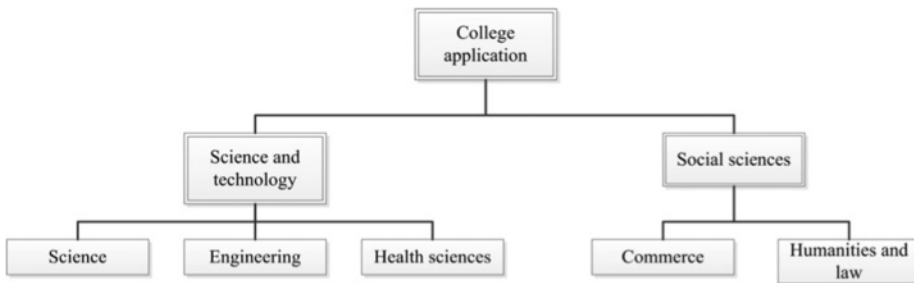
I use the two alternative specifications that are laid out in equations 5 and 6 to calculate aptitude-adjusted expected earnings for every applicant. To compute the first measure, I apply the formula in equation 5 to individualize the median income of each major by adjusting it with the probability of each applicant's success in college. I use the NBT scores to predict the probability of success. It is to be expected that each faculty has a specific requirement of skill sets determining success in that particular field. In order to determine which NBT module to use as a weight to calculate expected earnings in a given field, I run a probit regression of admission probability on the three types of NBT scores for the years prior to 2013. This is done under the assumption that university officials know the skill sets that are required to succeed in a given field and have already incorporated that information in the admission process. Based on the results of the probit regression (reported in Appendix A), math scores are used to adjust expected earnings in engineering, health sciences, and science, whereas academic literacy scores are used to adjust expected earnings in commerce and humanities and law.

In computing the second measure, I simply calculate the decile distribution of monthly income of each major and assign individuals to different deciles based on their aptitude. I assume the distribution of income for a certain major is aligned with the distribution of aptitude for that particular major. Accordingly, a person with a median NBT score can expect to earn the median income, whereas a person with a 95th-percentile NBT score can expect to fall in the top decile of the income distribution for that major. This is a simplistic measure that is intended to emphasize the potential impact of ability on labor market returns and inequality within the same major. As in the first measure, math scores are used to determine aptitude for engineering, health sciences, and science, whereas academic literacy scores are used in the cases of commerce and humanities and law. The last two rows in table 1 present descriptive statistics for the alternative measures of aptitude-adjusted expected earnings.

The next step concerns identification of the nesting structure that is supposed to govern the choice process of applicants. Intuitively, one can assume that applicants

**Table 2.** Correlation Between First and Second Choice Faculties

First Choice	Second Choice				
	Commerce	Engineering	Health Sciences	Science	Humanities and Law
Commerce	0.22	-0.09	-0.12	-0.08	0.02
Engineering	-0.04	0.32	-0.06	0.10	-0.22
Health sciences	-0.11	-0.02	0.27	0.21	-0.22
Science	-0.03	0.08	0.03	-0.11	-0.05
Humanities and law	-0.04	-0.23	-0.12	0.44	-0.15



**Figure 1.** Nesting structure

differentiate between the broad categories of science and technology on the one hand and social sciences on the other hand before picking specific faculties and departments within those categories. This assumption is corroborated by the clustering of skill sets required for admission along the science and technology versus social sciences divide. On top of all this, I am able to explicitly show the correlation of choices across faculties because there are data on the first and second choices of most applicants. Table 2 shows that an overwhelming number of applicants choose two majors in the same faculty as first and second choices. When applicants decide to choose a second major in another faculty, they mostly choose within the nests of social sciences and science and technology. These results justify the assumption that unobserved preference for specific majors is correlated within nests. Figure 1 displays the nesting structure that will be used to implement the nested logit estimation in the next section.

At this point, I can turn to the model specified in equation 8 to discuss the sources of empirical identification. Note that the main coefficients of interest in the vectors of parameters  $W_n$  and  $S_{n,j}$  relate to the effects of expected earnings and neighborhood characteristics. The fact that expected earnings are measured by the quantile income of contemporary workers can be used as an exclusion restriction to identify the effect of expected earnings on the choices of college applicants. However, the effect of quantile income is conditional on the probability of success or aptitude, which is measured by NBT results. The test scores of applicants in relevant modules are likely to influence major choice through channels other than the probability of success. Therefore,  $\mu_\beta$  in equation 6 may not be identified independently of  $p_{ij}$  unless there is exogenous variation in test results. The strategy I follow to identify  $\mu_\beta$  involves using the NSC scores in math, English, and life orientation to control for unobserved preference that

would otherwise be attributed to NBT results. For instance, a high score in NSC math is sufficient to predict whether an applicant would enjoy attending college in engineering. Once that effect is controlled for, the residual effect of NBT, which is designed to measure adaptability of prior knowledge to college curriculum, can be assumed to be affecting major choice through expected future earnings.

The coefficients of neighborhood effects are identified because they are estimated based on geographical variations. There is a valid concern that unobserved preference for educational outcomes could be correlated with the choice of residence. In other words, parents choose decent neighborhoods and good schools in the interest of better educational outcomes for their children. However, it is unlikely that parents would choose one neighborhood over the other because they want their children to study, say, engineering, in college. Even if they do, all they can consider, at high school level, is curriculum and quality, which are directly controlled for through high school science preparation and NSC grades for core subjects.

## 4. RESULTS

### Determinants of High School Curriculum Choice

Both theory and existing empirical evidence suggest that the choice of college major is conditional on high school preparation. More accurately, the choice of high school curriculum is often made in an anticipation of a certain college major and career path. This means an important part of the decision is already made in high school. However, individuals make two more nontrivial decisions at the point of college application. First, they can choose whether or not to switch the broad field they pursued in high school based on the information they have discovered about their own aptitude during high school. Second, whether or not they are switching fields, college applicants need to choose majors that are more specific than high school curriculums. I begin the presentation of results with the determinants of the choice of high school curriculum.

Table 3 presents the determinants of the number of high school science electives students take. I consider both individual-level and municipality-level correlates of curriculum choice. In column 1, all the explanatory variables are basic demographic and socioeconomic variables that are also included in the college major choice estimations (to be presented later). Both black and white students are found to take fewer science courses than Colored and Indians. The dummies for female and first-generation college are also negative, and public high schools are associated with more science courses. The positive coefficient of public high school is probably because those students from public high schools who apply to an elite university, such as UCT, self-select themselves on their inclination for technical subjects.

Considering that geographical location may have an impact on school quality and availability of some electives in such a historically segregated country as South Africa, I account for municipality-level effects by introducing 240 municipality dummies in column 1. Subsequently, in column 2, I examine what might be driving the municipality-level effects by estimating the correlation between key socioeconomic and political variables at the municipality level, and the municipality fixed-effects calculated based on column 1. Finally, column 3 controls for the municipality-level socioeconomic and political variables directly in estimating the determinants of high school curriculum.

**Table 3.** Determinants of Science Curriculum in High School

Dependent Variable	Number of Science Electives (1)	Municipality Fixed Effect (2)	Number of Science Electives (3)
Individual-level correlates			
Female dummy	-0.215*** (0.008)		-0.219*** (0.008)
Public school dummy	0.070*** (0.009)		0.061*** (0.009)
First-generation college dummy	-0.064*** (0.008)		-0.076*** (0.008)
Black dummy	-0.359*** (0.012)		-0.338*** (0.011)
White dummy	-0.086*** (0.012)		-0.080*** (0.011)
Municipality-level correlates			
Ratio of households above poverty line		-2.28*** (0.047)	-2.18*** (0.137)
Ratio of individuals with high school certificate		1.21*** (0.032)	1.14*** (0.100)
Percentage of ANC votes		.792*** (0.011)	.788*** (0.035)
Municipality fixed effect (240 municipality dummies)	Yes	No	No
R <sup>2</sup>	0.08	0.70	0.07
Number observations	11,644	234	11,644

Notes: Standard errors are given in parentheses. ANC = African National Congress.

\*\*\* Statistically significant at the 99% confidence level.

The results in columns 2 and 3 show that students applying from municipalities with higher levels of secondary school attainment are more prepared in science subjects. To the extent that high school completion is linked to availability of schooling resources, this result suggests that students from municipalities with better schooling resources come better prepared in science subjects. However, once educational attainment is controlled for, the coefficient of households above poverty line becomes strongly negative. This indicates that students from poorer municipalities apply to UCT probably because they are well prepared in high school science curriculum. Coupled with the positive coefficients of the public school variable in column 1, this finding points to the likelihood that students from poorer backgrounds self-select themselves to applying to UCT based on their preparation in high school science subjects. This result may have implications for the possibility of using improvements in high school science education in poorer areas as a means to reduce inequality through equitable access to elite higher education.

### The Effects of Expected Labor Market Returns

This section focuses on the role of expected labor market earnings in the choice of college major. The choice between science and technology on the one hand and social sciences on the other hand, which represents the first-level decision according to the nesting structure in figure 1, is specified as a function of high school curriculum or, alternatively, municipality-level variables influencing high school curriculum choice. The

choice of specific faculties, which is modeled as a second-level decision, is a function of major-specific expected earnings, family background, and demographic variables.

Table 4 presents the coefficient estimates of the random utility model of major choice. Because the magnitude of the coefficients in table 4 is not directly interpretable, average marginal effects of selected variables are presented in table 5 to provide an intuitive interpretation of the estimates. Columns 1 and 2 show that expected earnings, measured under alternative assumptions about ex ante distribution of earnings after graduation, exert a positive effect on the probability of choosing a given major. At the very least, the comparison between columns 1 and 2 demonstrates that the role of expected earnings is robust to alternative specifications with fundamentally different assumptions about the impact of ability on income distribution.

The effect of expected earnings, as measured by the more egalitarian indicator, disappears once high school curriculum is controlled for in column 4. In contrast, the meritocratic measure of expected earnings, which captures more information about post-college income differentials, continues to hold a strong effect on major choice independently of high school curriculum, as shown in column 5. It may be that the coefficient of the egalitarian measure is picking up the correlation between expected earnings and unobserved preferences, which is removed once high school curriculum is controlled for. The fact that this is not the case with the meritocratic measure suggests that expected earnings need to capture long-term income differentials to predict major choice net of unobserved preferences.

As columns 3 to 5 show, the number of science electives applicants took in high school increases the probability of choosing a major in college in science and technology. The choice of high school curriculum contains so much information about the aptitude of individuals for alternative fields with varying levels of financial desirability that it absorbs the explanatory power of major-specific median income. In other words, as far as median income can predict, applicants have already incorporated information on potential earnings when they decided on high school curriculum a few years earlier. Column 7 in table 5 displays that a one standard deviation increase in the number of high school science subjects raises the probability of choosing a major in science and technology by 0.22, on average. In order to account for the indirect effects of geographical factors on major choice, I control for municipality-level variables that are shown in the previous section to be correlated with the number of high school science electives applicants took. Accordingly, columns 6 and 7 show that applicants coming from poorer municipalities tend to choose majors in science and technology over majors in social sciences. This result indicates that prospective students from poorer areas apply to UCT mainly because they aspire to major in areas of science and technology.

The responsiveness of individuals to expected earnings might depend on the amount and quality of information they have about a number of factors, including the academic requirements of—and labor market returns to—different majors. As with many other things in South Africa, the information young people have about college and the labor market may be influenced significantly by their ethnic or racial background. In an attempt to disentangle the potential effect of race on shaping the responsiveness of applicants, the first two specifications in table 4 are repeated for the black and white subpopulations in columns 8 to 11. The results show that both measures of major-specific expected earnings have positive and statistically significant effects on

**Table 4.** Estimates of Random Utility Model of Major Choice with Market and Nonmarket Returns

Population	Full (1)	Full (2)	Full (3)	Full (4)	Full (5)	Full (6)	Full (7)	Black (8)	Black (9)	White (10)	White (11)
<i>In</i> expected earnings 1: egalitarian distribution	2.669*** (0.286)			0.203 (0.308)		1.893*** (0.312)		2.502*** (0.448)		4.117*** (0.575)	
<i>In</i> expected earnings 2: meritocratic distribution		0.375*** (0.035)			0.218*** (0.042)		0.297*** (0.037)		0.353*** (0.048)		0.498*** (0.083)
Level 1 Choice: Science and Technology (Reference: Social Sciences)			1.429*** (0.029)	1.252*** (0.032)	1.243*** (0.032)						
Number of high school science electives											
Ratio of households above poverty line in municipality						-2.591*** (0.717)	-2.440*** (0.707)				
Ratio of individuals with high school certificate in municipality						-0.857 (0.543)	-0.761 (0.542)				
Percentage of ANC votes in municipality						0.504*** (0.186)	0.470*** (0.185)				
Level 2 Choice: Specific Faculties (Reference: Humanities and Law)											
NSC math											
Commerce	0.052*** (0.002)	0.051*** (0.003)	0.609*** (0.088)	-0.000 (0.001)	0.097*** (0.017)	0.074*** (0.015)	0.054*** (0.012)	0.048*** (0.004)	0.047*** (0.004)	0.064*** (0.006)	0.062*** (0.006)
Engineering	0.048*** (0.003)	0.054*** (0.003)	0.565*** (0.087)	-0.023*** (0.004)	0.067*** (0.013)	0.083*** (0.014)	0.066*** (0.012)	0.049*** (0.005)	0.052*** (0.004)	0.054*** (0.008)	0.072*** (0.007)
Health sciences	0.033*** (0.002)	0.034*** (0.003)	0.202*** (0.033)	-0.001 (0.003)	0.056*** (0.011)	0.046*** (0.009)	0.036*** (0.008)	0.034*** (0.004)	0.034*** (0.004)	0.035*** (0.007)	0.045*** (0.007)
Science	0.031*** (0.003)	0.032*** (0.003)	0.186*** (0.035)	0.000 (0.004)	0.054*** (0.011)	0.042*** (0.010)	0.032*** (0.008)	0.031*** (0.005)	0.030*** (0.005)	0.033*** (0.008)	0.044*** (0.008)
Other case-specific controls: NSC math, NSC English, NSC life orientation, Public school dummy, First-generation college dummy, Black dummy, Female dummy			22.55	-1.166	0.591	2.151	1.651				
$\tau_n$ science and technology			38.26	-0.167	1.901	1.391	1.007				
$\tau_n$ social sciences			-13,169	-13,155	-15,385	-12,222	-13,114	-6,068	-6,057	-3,485	-3,493
Log-likelihood			9,267	9,267	9,267	9,239	9,239	4,312	4,312	2,467	2,467
Number of cases											

Notes: Standard errors are given in parentheses. Coefficients for seven additional controls listed in the table are not reported here. ANC = African National Congress.

\*\* Statistically significant at the 95% confidence level; \*\*\* statistically significant at the 99% confidence level.

Table 5. Average Marginal Effects<sup>a</sup>

Choice	Expected Earnings 1: Egalitarian Distribution (Own Effect)			Expected Earnings 2: Meritocratic Distribution (Own Effect)			Number of High School Science Electives <sup>f</sup>
	(1)	(2)	(3)	(4)	(5)	(6)	
	Black <sup>b</sup>	White <sup>c</sup>	White to Black ratio	Black <sup>d</sup>	White <sup>e</sup>	White to Black Ratio	
Science and technology							0.222*** (0.090)
Commerce	0.504*** (0.090)	0.913*** (0.128)	1.81	0.071*** (0.009)	0.110*** (0.018)	1.55	
Engineering	0.426*** (0.077)	0.617*** (0.089)	1.45	0.060*** (0.008)	0.074*** (0.013)	1.23	
Health sciences	0.561*** (0.101)	0.713*** (0.103)	1.27	0.079*** (0.011)	0.086*** (0.015)	1.09	
Science	0.229*** (0.042)	0.470*** (0.069)	2.05	0.032 (0.004)	0.056*** (0.009)	1.75	
Humanities and law	0.141*** (0.026)	0.465*** (0.068)	3.29	0.019*** (0.003)	0.056*** (0.009)	2.94	

Notes: <sup>a</sup>Standard deviations of the mean are given in parentheses. <sup>b</sup>Calculation based on column 8 in table 4. <sup>c</sup>Calculation based on column 10 in table 4. <sup>d</sup>Calculation based on column 9 in table 4. <sup>e</sup>Calculation based on column 11 in table 4. <sup>f</sup>Calculation based on column 3 in table 4.

\*\*\*Statistically significant at the 99% confidence level.

the probability of choosing a given major in both subpopulations. More importantly, columns 1 to 6 in table 5 present the average marginal effects of a one standard deviation increase in the natural logarithm of expected earnings in a certain major on the probability of choosing that major for the black and white subpopulations. Notably, white applicants are more responsive to changes in expected earnings across all majors and specifications of expected earnings. However, the gap in responsiveness between black and white applicants is the smallest for the more selective and financially rewarding majors, such as health sciences and engineering. In general, the meritocratic measure of expected earnings is shown to result in smaller gaps in responsiveness between black and white applicants. In other words, when we account for finer distinctions in potential income after graduation, black applicants appear more responsive to expected earnings than when we only take median income.

### Spatial Inequality and Neighborhood Effects

I have begun to account for the effect of spatial inequality on major choice through the use of municipality-level controls in the previous section. Nevertheless, spatial inequality in South Africa is much finer than differences at the municipality level. This section focuses on neighborhood effects down to the postcode and specific high school levels. First, I construct a measure of neighborhood-level admission trends for each of the 3,773 postcodes. This variable measures the ratio of students admitted to UCT to a given program from a single postcode out of the total number of students admitted from the same geographical area in the same year. This is calculated using admissions data between 2010 and 2012. I expect this variable to capture the influence of near-peer role models in a given neighborhood on the career decisions of college applicants. The notion of nearness in this context has both generational and spatial dimensions. The

**Table 6.** Average National Senior Certificate Scores in Student-School Quartiles

Quartiles of Students Within a School	Quartiles of Schools			
	1st	2nd	3rd	4th
1st	54.5	59.5	62.2	65.2
2nd	63.6	68.4	70.9	73.8
3rd	71.1	75.3	77.8	80.2 <sup>a</sup>
4th	78.3 <sup>b</sup>	82.9	85.5	87.1

Notes: <sup>a</sup>This cell represents a group of applicants from a highly competitive high school: Small Frog in a Big Pond. <sup>b</sup>This cell represents a group of applicants from a less competitive high school: Big Frog in a Small Pond.

neighborhood-level admission variable might be picking up a correlation between the unobserved preferences of applicants in the same neighborhood. Hence, in order to isolate the temporal near-peer effect, I correct for clustered errors at the postcode level.

In addition to influencing current decisions through past role models, the neighborhood effect might impact major choice by shaping the beliefs of individuals about their own ability. This hypothesis draws on established arguments about the “frog-pond” effect in the sociology of education literature (Davis 1966; Espenshade, Hale, and Chung 2005). The hypothesis predicts that a high-achieving student from a relatively less competitive school, that is, a big frog in a small pond (BFSP), tends to choose a more demanding major in college than a relatively low-achieving student from a more competitive school, that is, a small frog in a big pond (SFBP). In order to assign frog-pond identification to the population of applicants in the data, I create a pool of 1,557 high schools with at least five applicants to UCT between 2010 and 2013. Then I calculate a three-subject average NSC score for each school based on average scores in math, English, and life orientation. This makes it possible to allocate schools across a distribution of average grades divided into four quartiles. The next step is creating school-level quartiles of students using the same measurement as above. Table 6 displays the average grades for the sixteen student-school quartile combinations. The tightness of the distribution indicates that applicants to UCT already self-select themselves on their high school grades. I have identified one pair of BFSP and SFBP based on the criteria that the “big frogs” belong to the top quartile in their school even if their grades on average are lower than the “small frogs” in a more competitive school. The average score in the SFBP cell in table 6 is statistically greater than the average grade in the BFSP cell. I compare the effect of belonging to the BFSP cell as opposed to belonging to the SFBP cell on major choice, controlling for all other cell categories.

Table 7 presents the estimated coefficients of the random utility model with neighborhood and school effects described above. The results show that the impact of neighborhood-level admission trend on the choice for the respective major is highly significant under all specifications. The higher the number of students admitted to a certain field in the previous three years, the more likely it is for current applicants to choose the same field. The marginal effects obtained in table 8 show that the highest impact of near-peer role models is manifested in the choice of health sciences. A one standard deviation increase in the ratio of near-peers who were admitted to a health



**Table 7.** Estimates of Random Utility Model of Major Choice with Neighborhood and School Effects

Population	Full (1)	Full (2)	Full (3)	Full (4)	Full (5)	Full (6)
<i>ln</i> Expected earnings 1: Egalitarian distribution	0.386 (0.374)		0.798 (0.417)		1.231*** (0.374)	
<i>ln</i> Expected earnings 2: Meritocratic distribution		0.230*** (0.038)		0.234*** (0.044)		0.233*** (0.043)
Ratio of near-peers admitted to the same faculty	0.703*** (0.195)	0.724*** (0.172)			0.435** (0.220)	0.568*** (0.129)
Level 1 Choice: Science and Technology (Reference: Social Sciences)						
Number of high school science electives	1.244*** (0.041)	1.235*** (0.041)	1.249*** (0.042)	1.243*** (0.042)	1.238*** (0.042)	1.237*** (0.042)
Level 2 Choice: Specific Faculties (Reference: Humanities and Law)						
Dummy of Top-quartile NSC Score in a Bottom-quartile School (Reference: Dummy of 3rd Quartile NSC score in a Top-quartile School)						
Commerce			0.937** (0.490)	1.108** (0.542)	0.656 (0.439)	0.852** (0.428)
Engineering			0.987** (0.487)	0.972* (0.538)	0.494 (0.413)	0.553 (0.454)
Health sciences			1.059** (0.492)	1.284** (0.538)	0.862* (0.480)	1.078** (0.450)
Science			1.056** (0.492)	1.243** (0.548)	0.813* (0.477)	0.998** (0.474)
Other Case-specific Controls: NSC math, NSC English, NSC life orientation, Public school dummy, First-generation college dummy, Black dummy, Female dummy, 14 school-student quartile combinations						
$\tau_n$ Science and technology	0.907	0.986	0.116	0.487	0.596	0.851
$\tau_n$ Social sciences	1.613	1.847	0.855	1.032	0.556	0.727
Log-likelihood	-12,243	-12,227	-10,894	-10,893	-10,884	-10,875
Number of cases	9,267	9,267	8,334	8,334	8,334	8,334

Notes: Standard errors are given in parentheses. Coefficients for 21 additional controls listed in the table are not reported here. NSC = National Senior Certificate.

\*Statistically significant at the 90% confidence level; \*\* statistically significant at the 95% confidence level; \*\*\* statistically significant at the 99% confidence level.

**Table 8.** Average Marginal Effects of Ratio of Near-Peers Admitted to the Same Faculty<sup>a</sup>

Choice	(1) <sup>b</sup>	(2) <sup>c</sup>
Commerce	0.104*** (0.037)	0.104*** (0.037)
Engineering	0.115*** (0.052)	0.110*** (0.049)
Health sciences	0.147*** (0.041)	0.143*** (0.040)
Science	0.071*** (0.026)	0.067*** (0.025)
Humanities and law	0.040*** (0.032)	0.037*** (0.030)

Notes: <sup>a</sup>Standard deviations of the mean are given in parenthesis. <sup>b</sup>Calculation based on column 1 in table 7. <sup>c</sup>Calculation based on column 2 in table 7.

\*\*\*Statistically significant at the 99% confidence level.

sciences faculty during the last three years increases the probability of choosing health sciences by over 14 percent. On the contrary, the same increase in the case of humanities and law results in no more than a 4 percent rise in the probability of choosing the same field.

Columns 3 to 6 in table 7 include the frog-pond indicators generated based on table 6. The frog-pond effect is mainly about the self-evaluation of students in high school in comparison with their schoolmates. Therefore, the effect is more accurately measured after controlling for choices in high school, such as high school curriculum. The results show that there is an indication of a frog-pond effect in the choice of commerce, health sciences, and science under most specifications. For engineering, the effect is limited to specifications excluding admission trends of near-peer role models. These results are unlikely to be spurious because, out of twelve student-school quartile combinations that are characterized by average NSC scores lower than the SFBP cell, the only category that has a positive and significant effect on any choice is the BFSP cell. It might be interesting to put these results in perspective with the effect of near-peer role models. The path-dependent effect of neighborhood-level admission trends may suggest that applicants from less competitive schools might end up choosing low return majors. The frog-pond effect adds a twist to that story. After all, the career destiny of some students from less competitive and, presumably, underresourced schools might be improved by the fact that they have maintained relatively high self-esteem coming out of high school.

## 5. CONCLUSION

This paper has set out to examine the determinants of college major choice against the backdrop of inter-group and spatial inequalities. I have used the extensive set of academic, geographical, and socioeconomic information contained in the admissions application database of the University of Cape Town to estimate random utility models of major choice. In line with the predictions of the lifecycle model of career choice, the choice of high school curriculum is shown to be crucial for the choice of college major. However, expected earnings maintain a positive impact on major choice independently of high school background when the ex ante distribution of earnings captures the full range of between-major and within-major income differentials. At a neighborhood level, the influence of near-peer role models on the choices of college applicants is found to be large and significant.

The dynamics of major choice at a selective institution such as UCT is likely to have significant implications in the long run for social and economic transformation. Potential inefficiencies in the allocation of talent that might be caused by persistent inequalities will hamper innovation and mass flourishing à la Edmund Phelps (Phelps 2014). Although the gap between white and black students in response to potential differentials in expected earnings remains, it is encouraging that the magnitude is significantly smaller in the case of high-return majors, such as health sciences. Moreover, good preparation in high school science electives seems to embolden applicants from poorer and far-off municipalities to apply to UCT, hoping to major in science and technology fields. Considering the importance of high school curriculum and neighborhood effects, policy measures that will improve the availability of science education at

the high school level or account for the effect of near-peer role models in college admissions may go a long way in terms of moderating income inequality in South Africa.

### ACKNOWLEDGMENTS

I would like to thank the Information and Communication Technology Services of the University of Cape Town for allowing me access to the admissions database. I am grateful to Kende Kefale for facilitating access to the database. I thank Max Price and two anonymous referees for comments and suggestions. I thank Chris Rooney and Callee Davis for excellent research assistance. Financial support for this research has been provided by Economic Research Southern Africa. This paper was completed during my time at Princeton University.

### REFERENCES

- Altonji, Joseph G. 1993. The demand for and return to education when education outcomes are uncertain. *Journal of Labor Economics* 11(1): 48–83.
- Altonji, Joseph G., Erica Blom, and Costas Meghir. 2012. Heterogeneity in human capital investments: High school curriculum, college major, and careers. *Annual Review of Economics* 4(1): 185–223.
- Arcidiacono, Peter. 2004. Ability sorting and the returns to college major. *Journal of Econometrics* 121(1–2): 343–375.
- Bedasso, Biniam E. 2015. Educated bandits: Endogenous property rights and intra-elite distribution of human capital. *Economics & Politics* 27(3): 404–432.
- Bénabou, Roland. 1996. Equity and efficiency in human capital investment: The local connection. *Review of Economic Studies* 63(2): 237–264.
- Braddock II, Jomills Henry. 1980. The perpetuation of segregation across levels of education. *Sociology of Education* 53(3): 178–186.
- Card, David, and Jesse Rothstein. 2007. Racial segregation and the black-white test score gap. *Journal of Public Economics* 91(11–12): 2158–2184.
- Davis, James A. 1966. The campus as a frog pond: An application of the theory of relative deprivation to career decisions of college men. *American Journal of Sociology* 72(1): 17–31.
- Daymont, Thomas N., and Paul J. Andrisani. 1984. Job preferences, college major, and the gender gap in earnings. *Journal of Human Resources* 19(3): 408–428.
- Espenshade, Thomas J., Lauren E. Hale, and Chang Y. Chung. 2005. The frog pond revisited: High school academic context, class rank, and elite college admission. *Sociology of Education* 78(4): 269–293.
- Goldsmith, Pat Rubio. 2009. Schools or neighborhoods or both? Race and ethnic segregation and educational attainment. *Social Forces* 87(4): 1913–1941.
- Grogger, Jeff, and Eric Eide. 1995. Changes in college skills and the rise in the college wage premium. *Journal of Human Resources* 30(2): 280–310.
- Keane, Michael P., and Kenneth I. Wolpin. 1999. The career decisions of young men. *Journal of Political Economy* 105(3): 473–522.
- Montmarquette, Claude, Kathy Cannings, and Sophie Mahseredjian. 2002. How do young people choose college majors? *Economics of Education Review* 21(6): 543–556.

Murphy, Kevin M., Andrei Shleifer, and Robert W. Vishny. 1991. The allocation talent: Implications for growth. *Quarterly Journal of Economics* 106(2): 503–530.

Phelps, Edmund. 2014. *Mass flourishing: How grassroots innovation created jobs, challenge and change*. Princeton, NJ: Princeton University Press.

Van der Berg, Servaas. 2007. Apartheid's enduring legacy: Inequalities in education. *Journal of African Economies* 16(5): 849–880.

Weinberger, Catherine. 1998. Race and gender wage gaps in the market for recent college graduates. *Industrial Relations* 37(1): 67–84.

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

## APPENDIX A: PROBIT REGRESSION RESULTS

**Table A.1.** Probit Estimates of Admission Probability as a Function of National Benchmark Test Scores (Pooled Population of Applicants from 2010 to 2012)

	Commerce	Engineering	Health Sciences	Science	Humanities and Law
Academic Literacy score	0.026*** (0.002)	0.010*** (0.002)	0.017*** (0.003)	0.002 (0.003)	0.023*** (0.003)
Quantitative Literacy score	0.004** (.002)	−0.001 (0.002)	−0.003 (0.002)	0.004 (0.003)	−0.001 (0.003)
Mathematics Score	0.024*** (0.002)	0.040*** (0.001)	0.020*** (0.002)	0.028*** (0.002)	0.013 (0.003)
Pseudo R <sup>2</sup>	0.14	0.20	0.10	0.11	0.08
Log likelihood	−3,516	−2,174	−1,732	−1,161	−1,094
Number of observations	6,452	5,050	5,970	2,144	2,008

Notes: Standard errors are given in parentheses.

\*\* Statistically significant at the 95% confidence level; \*\*\* statistically significant at the 99% confidence level.

## APPENDIX B: DATA SOURCES

- (1) Admission application data: The database of the Information and Communication Technology Services of the University of Cape Town.
- (2) Earnings Data: Quarterly labor force survey, Statistics South Africa.