

Yao, Kan

Article

Heterogeneous skill distribution and college major: Evidence from PIAAC

Journal of Applied Economics

Provided in Cooperation with:

University of CEMA, Buenos Aires

Suggested Citation: Yao, Kan (2019) : Heterogeneous skill distribution and college major: Evidence from PIAAC, *Journal of Applied Economics*, ISSN 1667-6726, Taylor & Francis, Abingdon, Vol. 22, Iss. 1, pp. 504-526,
<https://doi.org/10.1080/15140326.2019.1665310>

This Version is available at:

<https://hdl.handle.net/10419/314071>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>



Heterogeneous skill distribution and college major: evidence from PIAAC

Kan Yao

To cite this article: Kan Yao (2019) Heterogeneous skill distribution and college major: evidence from PIAAC, *Journal of Applied Economics*, 22:1, 504-526, DOI: [10.1080/15140326.2019.1665310](https://doi.org/10.1080/15140326.2019.1665310)

To link to this article: <https://doi.org/10.1080/15140326.2019.1665310>



© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 15 Oct 2019.



Submit your article to this journal [↗](#)



Article views: 1344



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 4 View citing articles [↗](#)

ARTICLE



Heterogeneous skill distribution and college major: evidence from PIAAC

Kan Yao

Department of Economics, University of California, Los Angeles (UCLA), Los Angeles, CA, USA

ABSTRACT

A large literature documents the uneven distribution of labor market outcomes across majors. Students in STEM (science, technology, engineering, and mathematics) can earn more than their peers. This earnings gap can be attributed not only to the differential educational resources investment but also to heterogeneous distribution of initial cognitive skills across majors. I benefit from the rich data from the Programme for the International Assessment of Adult Competencies to examine this earnings gap in the United States and the United Kingdom. Based on my findings, this paper establishes new facts that add to the understanding of how college field premiums are generated. I show that a sizable portion of the return to majors is due to self-selection and up to two-fifths of the field premiums can be explained by basic cognitive skills. Despite the qualitatively similar impacts of numeracy and literacy skills on choosing college field of study, the pricing of numeracy is much higher than literacy in the labor market.

ARTICLE HISTORY

Received 18 November 2017

Accepted 13 August 2019

KEYWORDS

Wage premiums; cognitive skills; field of study

1. Introduction

A large amount of literature in economics discusses the aggregate college premium over non-college, but little focuses on the substantial earnings gap across college majors. Carnevale, Cheah, and Hanson (2015) use the 2009 American Community Survey (ACS) to investigate the inter-major earning difference. They find that students in science, technology, engineering, and mathematics (STEM) earn on average 65% more than those in teaching and service majors. STEM students also enjoy a higher earnings growth rate over the course of a career than those in non-STEM majors. The highest-earning majors, besides STEM majors, also include pharmacy pharmaceutical sciences and administration. In fact, some economists have reported that the difference in returns across college majors even overshadows the college earnings premium over non-college (Altonji, Blom, & Meghir, 2012). Altonji, Kahn, and Speer (2014) point out that the earnings gap across college majors has been widened through years and it is likely to become one primary source of inequality in developed countries.

College earnings gap reflects not only the differential investments into specific educational resources, but also the heterogeneous skills of individuals across majors

CONTACT Kan Yao  yaokan@ucla.edu

© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

(Arcidiacono, 2004; Paglin & Rufolo, 1990; van de Werfhorst, 2002). Difference in prospective college students' skills affects their choices of major, leading to the heterogeneous distributions of skills across majors. The heterogeneity in skills, combined with difference in college course contents, drives variation in the choices of professional career paths across majors, thereby altering their earnings (Altonji et al., 2014). This process generates the field premium.

The observed correlation between earnings gap and majors is not causal due to the underlying self-selection in schooling decisions and can be partly explained by the heterogeneity in skills. A key problem of explaining the current earnings gap across majors, however, is that dataset with information of earnings and college majors does not have precise measurement of one's cognitive skills, the most studied ability in the economics literature. One approach literature adopts is to use the standardized achievement test scores as a proxy to one's initial ability prior to college. For instance, Paglin and Rufolo (1990) define mathematics ability in terms of quantitative scores in Scholastic Assessment Test (SAT) and Graduate Record Examinations (GRE) and disentangle the returns to the use of scarce mathematics ability from the returns to college major. Arcidiacono (2004) models the ability sorting across college majors using SAT verbal and math scores as indicators of one's verbal and math abilities. However, performance in the achievement test, such as SAT, is also significantly affected by efforts to prepare for the test and character skills, such as self-control (Duckworth, Quinn, & Tsukayama, 2012). Heckman and Kautz (2014) argue that character skills can explain the variance in achievement test scores above and beyond the variance cognitive skills do. Thus, when we use SAT scores as a proxy to one's cognitive ability, it would overstate the relative importance of the cognition by including the returns to noncognitive skills as well.

To investigate the impact of cognitive skills on college earnings gap across majors, I explore the United States and the United Kingdom data of the 2012 Programme for the International Assessment of Adult Competencies (PIAAC), which has two advantages. It contains the sophisticated measurements of adults' initial cognitive skills, including numeracy and literacy skills. It also collects rich background information including respondents' educational qualifications and their labor market outcomes.

I first examine the relation between college major choices and two unidimensional skills – numeracy and literacy. A multinomial logit model is used. It turns out that higher basic cognitive skills are associated with less probability of choosing subjects such as education, humanities and arts but increase the likelihood of choosing STEM majors. This result complies with literature on ability sorting across majors (Arcidiacono, 2004; Arcidiacono, Hotz, & Kang, 2012). Moreover, numeracy is a stronger predictor of one's field choices in STEM. For instance, the average increase in possibility of choosing engineering, manufacturing and construction is as three times associated with one-standard-deviation increase in numeracy as literacy.

I also formulate a multivariate model to disentangle the impact of cognitive skills from the observed field premiums. I use the model to estimate to what extent the variations in skills can explain field earnings gap. The estimates indicate that up to two-fifths of field premiums are explained by the return to numeracy skills. Numeracy is much more valued in the labor market compared to literacy. However, field premiums are not merely generated via sorting

high-skill people into more lucrative fields and awarding for their human capital. Large premiums do exist, for example, for the field of health and welfare.

The remainder of this paper is organized as follows: [Section 2](#) provides information on the data set that I use in this paper, including the statistical description. [Section 3](#) presents the inference and empirical results. [Section 4](#) discusses results and concludes.

2. Data

The PIAAC is a large-scale comparative survey that was developed and implemented by the Organization for Economic Cooperation and Development (OECD). Thirty-three countries participated in this study from 2011 to 2015¹, and PIAAC released the updated Main Study database in 2016. In this paper, I use data from the United States and the United Kingdom for my analysis, since these two countries share some similarities in cultural background and educational system. I also benefit from the US National Supplement provided by National Center for Education Statistics (NCES).

PIAAC is designed to measure key cognitive skills and competencies that underlie one's labor market outcomes and social success. It attempts to obtain a nationally representative sample of the whole population aged from 16 to 65 in each participating country that is proportional to the population across the country².

To study the higher education field premiums, I only focus on native respondents with college degree or above who are active in the labor market as workers³. Since the unemployment rate is merely 4.8% in the US and 2.9% in the UK among my data sample, it is of second order of concern. I regard unemployment as due to typical labor market frictions and therefore treat it as random missing in monthly earnings. I further exclude those who are employed but report zero income as outliers. [Table 1](#) presents basic statistic description.

2.1. Fields of study

The main variable of interest is the field of study. Classified based on the 2011 International Standard Classification of Education (ISCED), it has nine categories as follows: general programs, education, humanities and arts, social science, business and law, science and mathematics, engineering, manufacturing and construction, agriculture and veterinary, health and welfare, services⁴. In the US National Supplement, the

¹There are two rounds for the first cycle of PIAAC. Twenty-four countries participated in the Round 1 of the PIAAC with data collection taking place from 1 August 2011 to 31 March 2012 in most countries. Nine countries took part in the Round 2 of the assessment with data collection taking place from April 2014 to end-March 2015. The first-round countries are: Australia, Austria, Belgium, Canada, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Ireland, Italy, Japan, Korea, Netherlands, Norway, Poland, Russian Federation, Slovak Republic, Spain, Sweden, the United Kingdom, and the United States. The second-round countries are: Chile, Greece, Indonesia, Israel, Lithuania, New Zealand, Singapore, Slovenia, and Turkey.

²The respondents' selection criteria has four stage in the United States and the United Kingdom. The main stratification variables for the United States include region, metro area classification, race/ethnicity, income, percentage of the population that is foreign born and the sample is further sorted by geographic location. The main stratification variables for the United Kingdom include region, percentage living in social housing and the sample is further sorted by percentage of White British; postcode and address number; addresses, and first name. A more detailed description of sample units and sample selection methods is in PIAAC Technical Report [Section 4](#) (OECD, 2013).

³I exclude full-time students from my data sample to obtain a more precise estimation.

⁴A detailed classification of field education and training is Appendix B.

Table 1. Sample descriptive statistics.

	Pooled	U.S.	U.K.
<i>Gender</i>	2	3	4
Female	57.9	57.0	58.4
Male	42.1	43.0	41.6
<i>Age</i>			
Entry Age	33.0	30.8	34.5
Prime Age	50.5	48.0	52.1
Exit Age	16.5	21.3	13.4
<i>Number Living in Household</i>			
Single	23.1	26.0	21.4
Two	30.7	31.3	30.3
Three	17.9	16.8	18.5
Four or Above	28.3	26.0	29.8
<i>Work Experience</i>			
No More than 10 Years	23.3	21.1	24.6
11 to 20 Years	32.0	26.6	35.5
21 to 30 Years	24.0	23.6	24.2
No Less than 31 Years	20.7	28.8	15.7
No. Obs.	2485	959	1526

This table presents the basic statistic description. Column 2 presents the percentage in the whole data sample; Column 3 presents the percentage conditional on US data; Column 4 presents the percentage conditional on UK data. Information of race and region is only available for US from its National Supplement.

Table 2. Breakdown of respondents by field of study (in percentage).

	Pooled	US	UK
<i>STEM Majors</i>	2	3	4
Science and Mathematics	14.8	12.3	16.3
Engineering, Manufacturing and Construction	8.8	5.4	10.9
<i>Non-STEM Majors</i>			
General Programs	2.5	2.6	2.5
Education	11.9	15.4	9.6
Humanities and Arts	17.5	14.6	19.4
Social Science, Business and Law	31.4	33.7	30.0
Agriculture and Veterinary	0.6	1.0	0.4
Health and Welfare	11.1	12.3	10.3
Services	1.3	2.6	0.5
No. Obs.	2485	959	1526

This table presents the breakdown of respondents by field of study. Column 2 presents the percentage in the whole data sample; Column 3 presents the percentage conditional on US data; Column 4 presents the percentage conditional on UK data.

sixth and seventh fields are combined as engineering, manufacturing, and agriculture. In my analysis, I adopt the nine categories defined in the PIAAC Main Study.

Table 2 shows the proportion of each field of study for each country. The assignment of fields of study is based on one's highest qualification. Around one-quarter of people choose STEM majors in both countries⁵. A majority of people are in the field of social science, business and law. Since the sample size of those who are in the field of "agriculture and veterinary", I will exclude them in my estimations to avoid unnecessary bias.

⁵STEM majors refer to respondents in the fields of science, mathematics and computing or engineering, manufacturing and construction.

2.2. Earnings data

My main labor market outcome measurement is the monthly earnings, purchasing power parity (PPP) adjusted in dollar. It is a continuous variable obtained from the UK Public Use File and the US National Supplement.⁶ Among the available employed respondents, up to 7% of the income data are not reported and I exclude them in my analysis in Section 3. This attrition may correspond to the potential existence of nonresponse bias. In Section 3.3, I perform a robustness check by using Inverse Probability Weighting (IPW) method. The results indicate that nonresponse bias is not of major concerns.

Table 3 presents the income information associated with each field. On average, the UK respondents on average have less purchasing power than those in the US. The highest average monthly earning is observed in the field of social science, business, and law. Since ISCED compresses a variety of majors with difference earning potentials within this field, the high average income may result from a greater proportion of high-earning-potential majors such as economics⁷ (Altonji et al., 2012). The field of agriculture and veterinary is a special case. It has very few people but provides the highest average earnings. It might be due to that these people inherit or work for familial

Table 3. Earnings statistics by field of study.

	Pooled	U.S.	U.K.
<i>STEM Majors</i>	2	3	4
Science and Mathematics	5.74 (6.0)	6.78 (5.6)	5.25 (6.1)
Engineering, Manufacturing and Construction	5.46 (3.9)	7.12 (4.1)	4.95 (3.8)
<i>Non-STEM Majors</i>			
General Programs	5.21 (6.4)	5.32 (3.7)	5.13 (7.8)
Education	4.74 (8.0)	5.34 (9.4)	4.14 (6.3)
Humanities and Arts	4.07 (3.6)	4.92 (3.9)	3.67 (3.4)
Social Science, Business and Law	5.80 (6.4)	7.29 (7.8)	4.74 (5.0)
Agriculture and Veterinary	6.93 (5.2)	7.34 (4.7)	6.26 (6.4)
Health and Welfare	5.25 (4.6)	6.07 (4.9)	4.63 (4.3)
Services	5.51 (3.8)	6.10 (4.1)	3.69 (1.6)
Total	5.26 (5.8)	6.34 (6.8)	4.58 (5.0)
No. Obs.	2485	959	1526

This table presents the earnings data by field of study. “Income” refers to monthly earnings in thousand dollars, PPP adjusted. Column 2 presents the earnings mean and standard errors in the whole data sample; Column 3 presents the earnings mean and standard errors conditional on U.S. data; Column 4 presents the earnings mean and standard errors conditional on U.K. data. Standard errors are shown in brackets.

⁶The continuous income measurement is classified in the US Public Use File.

⁷Since the data does not contain information about what specific majors respondents are in, we cannot know to what extent the intra-field earning difference has been.

pasture after graduation. Another explanation would link to neoclassic economic theory that the scarcity of veterinarian supply may increase their profits.

Consistent with most literature in labor economics, i.e., Mincer (1974), I will use the logarithm-transformed income as my dependent variable. In both countries, STEM majors have the highest log income.

2.3. Skills

PIAAC has three well-assessed skills in the domains of literacy, numeracy, and problem-solving in the technology-rich environment. They are measured on a 500-point scale and PIAAC uses adaptive testing and a combination of an Item Response Theory (IRT) model and a latent regression model to generate 10 plausible values for each skill measurement.⁸ They are defined as⁹:

- Literacy: “understanding, evaluating, using and engaging with written texts to participate in society, to achieve one’s goals, and to develop one’s knowledge and potential.”
- Numeracy: “the ability to access, use interpret and communicate mathematical information and ideas, in order to engage in and manage the mathematical demands of a range of situations in adult life.”
- Problem-solving in technology-rich environments: “using digital technology, communication tools and networks to acquire and evaluate information, communicate with others and perform practical tasks.”

There is relatively large portion of missing in problem-solving skills in the technology-rich environment, because the problem-solving skill can only be assessed via computer and the corresponding questionnaire was not carried out for those who did not take PC version.¹⁰ Therefore, I will exclude problem-solving skills in the technology-rich environment from my analysis in the rest of paper.

Figure 1 presents the sample decomposition based on numeracy and literacy skills, respectively.¹¹ I divide people into two groups: those whose skill scores are above the overall average and those who have below average skill scores. The figures show a sizable variation in skills by field of study. STEM majors have most people who have above average numeracy skills (67.6% for engineering, manufacturing, and construction and 66.8% for science and mathematics). The field of agriculture and veterinary has the highest proportion of people with above average literacy skills, followed by the STEM majors. On the contrary, more skill scores (both numeracy and literacy) in the field of services are below the overall average.

This may cast some doubt on the interpretation of the decomposition of earnings gap across fields of study, since cognitive skills are measured when individuals are

⁸A more detailed description of how each cognitive skill score is defined and generated is in PIAAC Technical Report Section 5 (OECD, 2013).

⁹Consistent with other literature on PIAAC, i.e., Allen and Van der Velden (2001); Hanushek et al. (2015), I will use the first plausible value as an indicator of each skill.

¹⁰The standard cognitive assessment was mainly carried out on computers but literacy and numeracy components can also be done via pencil-and-paper test for respondents who had extremely limited IT experience.

¹¹See Appendix A.2 for details on numeracy and literacy skill distributions.

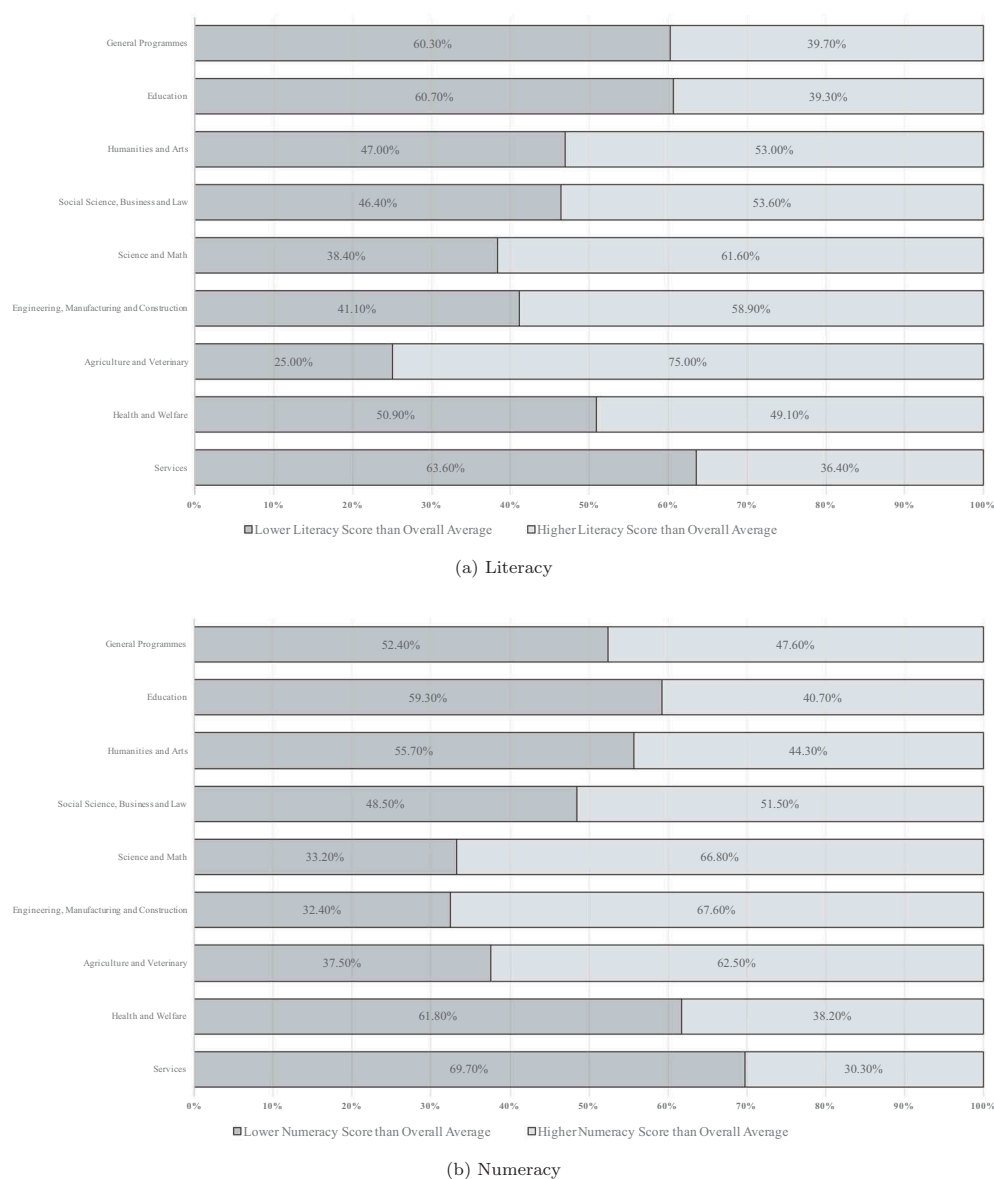


Figure 1. Skill score composition by field of study.

already in the labor market instead of prior to attending college. Namely, it may reflect the value of human capital one acquire throughout educational investment and work experience instead of innate skills. However, the way how PIAAC assessed skills help to reject such conjecture, because the assessment tasks were similar to the way people might encounter in different “real life contexts” and different from solving “school-like problems” (PIAAC Numeracy Expert Group, 2009). Particularly, no college-level knowledge was required and tested.¹² That is, the skill scores were unlikely to be affected by one’s college educational process and his/her college major choice.

Moreover, a large literature in economics, psychology and education indicates that one's intelligence scores become stable after around age 10 and school levels have little effect on reducing this score gap among different socioeconomic groups (Heckman, Carneiro, & Cunha, 2004; Heckman, Stixrud, & Urzúa, 2006; Hopkins & Bracht, 1975). The stability of one's intelligence allows us to use skill scores measured when individuals are already in the labor market as if those were measured before attending college. Thus, the PIAAC-measured skills are likely to reflect one's attribute prior to college instead of specific human capital investment acquired through education and I also use them to control for self-selection in the next section.

3. Analysis and results

The observed heterogeneity in skill distributions across fields of study motivates the investigation of the interplay between one's cognitive skills and field preference. For analytical purpose, I standardize each cognitive skill by age and countries. They have zero-mean and one-standard-deviation in each age category within each country.¹³

3.1. Choices of field of study

Equation (1) describes the basic choice model for college field of study. U_{ij} is the utility of choosing field of study $j \in \mathcal{J}$ for individual $i \in \mathcal{I}$, where \mathcal{I} denotes the sample space and \mathcal{J} denotes all possible choices of major. Individual i is assumed to compare the expected utilities associated with each field among \mathcal{J} alternatives. Denote η_{ij} and ϵ_{ij} to be systematic and random component of one's utility. Each component is individual-and-field specific. The systematic component, η_{ij} , is assumed to be consisted of two parts. The first term, γ_j , is the utility of field itself. It is the same across all individuals and is independent of one's skills. The other term, $\theta_j \cdot \mathbf{S}_i$, is related to the return to i 's own skills when applying his or her skills, \mathbf{S}_i , to field j . That is

$$\begin{aligned} U_{ij} &= \eta_{ij} + \epsilon_{ij} \\ \eta_{ij} &= \gamma_j + \theta_j \cdot \mathbf{S}_i \end{aligned} \quad (1)$$

Literature, e.g., Zafar (2013) and Turner and Bowen (1999), discusses how male and female value college fields differently. In recognition of this gender difference, I control for the gender effect in my model. I also allow for country difference in field preference. Now the model reads

$$\eta_{ij} = \gamma_j + \theta_j \cdot \mathbf{S}_i + \sum_{g=1}^{G-1} \varphi_g \cdot 1\{G_i = g\} + \sum_{n=1}^{N-1} \varphi_n \cdot 1\{N_i = n\} \quad (2)$$

¹²Though the whole task pool of PIAAC is not publicly accessible, its own report and some sample questions could help us understand the complexity of, for example, mathematical information required. For instance, one numeracy sample question is to ask "which period was there a decline in the number of births" based on a line chart, which presents every 10 years, of "the number of births in the United States from 1957 to 2007" (Gal & Tout, 2014; OECD, 2009). Even at its hardest level, only topics such as rational numbers, dimensions and probability are examined for numeracy assessment (PIAAC Numeracy Expert Group, 2009). Therefore, it is safe to say that PIAAC does not require any college-level knowledge for its skill assessment.

¹³Age categories for my data sample are 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59 and 60-65.

where $1\{G_i = g\}$ is the binary indicator of individual i 's gender G_i is g , and $1\{N_i = n\}$ indicates if i is from country n .

Assume that individual i choose field j to maximize his or her utility. Let J_i indicates i 's field choice such that

$$J_i = j \Leftrightarrow U_{ij} = \max_{\tilde{j} \in \mathcal{J}} U_{i\tilde{j}} \quad (3)$$

To simplify the estimation, I use the conditional logit model of McFadden (1973) to get empirical estimates of the parameters of Equation (2).¹⁴ Therefore, the probability that individual i chooses field j , π_{ij} , can be stated as in Equation (4).

$$\pi_{ij} = \Pr(J_i = j) = \frac{\exp(\eta_{ij})}{\sum_{\tilde{j}=1}^{\mathcal{J}} \exp(\eta_{i\tilde{j}})} \quad (4)$$

where η_{ij} contains skills, gender, and country effects, as defined in Equation (2).

In Table 4, I estimate multinomial logit models and each model is dedicated to a unidimensional skill, i.e., numeracy and literacy. Column 2 and 4 presents coefficients in the relative-risk ratios. For example, 57.1% (14.2%) increase in the relative probability of choosing science and mathematics over humanities and arts is associated with one-standard-deviation increase in one's numeracy (literacy) index¹⁵.

Figure 2 describes the predicted possibility of choosing each field against numeracy and literacy skills, respectively. One increase in the x -axis stands for one-standard-deviation increase in the skill scores. We observe that the slope of marginal effect of increasing skill on the probability of choosing a college field of study is largely similar for both skills. That is, an increase in either skill increases the likelihood of choosing STEM majors while shifts people away from other fields such as education, humanities, and arts. There is little effect of either skill on choosing general programs. The one exception to the observed similarity is for women regarding humanities and arts. Women are less likely to choose humanities and arts with an increase in numeracy skills, while the possibility increases slightly if their literacy skills increase.

The striking similarity of these graphs across skills strongly suggests that both skills simply present different facets of the same latent cognitive trait. Since cognitive skill development is self-productive and cross-fertilizing, we may expect the observation of such latent cognitive trait (Cunha & Heckman, 2009).

Both figures present a non-monotonic relationship with skills in the field social science, business and law. Since PIAAC uses ISCED to classify fields of study, we cannot distinguish majors like anthropology, economics, business, or law within the same field termed as "social science, business and law." It might be the case that with increase in numeracy or literacy, people may move away from other social science, e.g., anthropology, but at the same time people are more likely to major in economics (Turner & Bowen, 1999). The net change in possibility of choosing social science,

¹⁴The logit model requires some normalization assumptions to be made. Specifically, I assume "humanities and arts" to be the "universal benchmark" choice. Also, i 's unobserved preference for a particular major j to follow standard Type I extreme value distribution can be obtained by imposing some mild conditions on the distribution of the random component (McFadden, 1973).

¹⁵I use "numeracy index" ("literacy index") in the rest of my paper to refer to the standardized numeracy (literacy) skill scores measured by PIAAC.

Table 4. Multinomial logit estimates and average marginal effect.

	Model 1		Model 2	
	Numeracy Index		Literacy Index	
	Coefficient	AME	Coefficient	AME
<i>STEM Majors</i>	2	3	4	5
Science and Mathematics	1.571*** (0.12)	0.041*** (0.01)	1.142+ (0.08)	0.022** (0.01)
Engineering, Manufacturing and Construction	1.760*** (0.16)	0.030*** (0.01)	1.122 (0.10)	0.010+ (0.01)
<i>Non-STEM Majors</i>				
General Programs	0.829 (0.12)	−0.008* (0.00)	0.672** (0.09)	−0.009** (0.00)
Education	0.958 (0.08)	−0.018** (0.01)	0.784** (0.06)	−0.023*** (0.01)
Humanities and Arts	–	−0.022** (0.01)	–	0.005 (0.01)
Social Science, Business and Law	1.150* (0.07)	0.001 (0.01)	0.995 (0.06)	0.008 (0.01)
Health and Welfare	0.933 (0.08)	−0.019** (0.01)	0.900 (0.07)	−0.007 (0.01)
Services	0.683* (0.13)	−0.006* (0.00)	0.622** (0.11)	−0.006* (0.00)

+ $p < .1$, * $p < .05$, ** $p < .01$, *** $p < .001$
This table presents estimates based on two multinomial logit models. The data sample consists of $N = 2469$, using both US and UK data samples. Individuals in agriculture and veterinary major are excluded due to limited number of observations. Columns 2–3 present the coefficients in the relative-risk ratios and average marginal effect of numeracy index in the model of Equation (2), in which S_i is a unidimensional numeracy index. The model reads as follows,

$$U_{ij} = \nu_j + \theta_j^N \cdot S_i^N + \sum_{g=1}^{G-1} \varphi_g \cdot 1\{G_i = g\} + \sum_{n=1}^{N-1} \varphi_n \cdot 1\{N_i = n\} + \epsilon_{ij}$$

Columns 4–5 present the coefficients in the relative-risk ratios and average marginal effect of literacy index in the model of Equation (2), in which S_i is a unidimensional literacy index. The model reads as follows,

$$U_{ij} = \nu_j + \theta_j^L \cdot S_i^L + \sum_{g=1}^{G-1} \varphi_g \cdot 1\{G_i = g\} + \sum_{n=1}^{N-1} \varphi_n \cdot 1\{N_i = n\} + \epsilon_{ij}$$

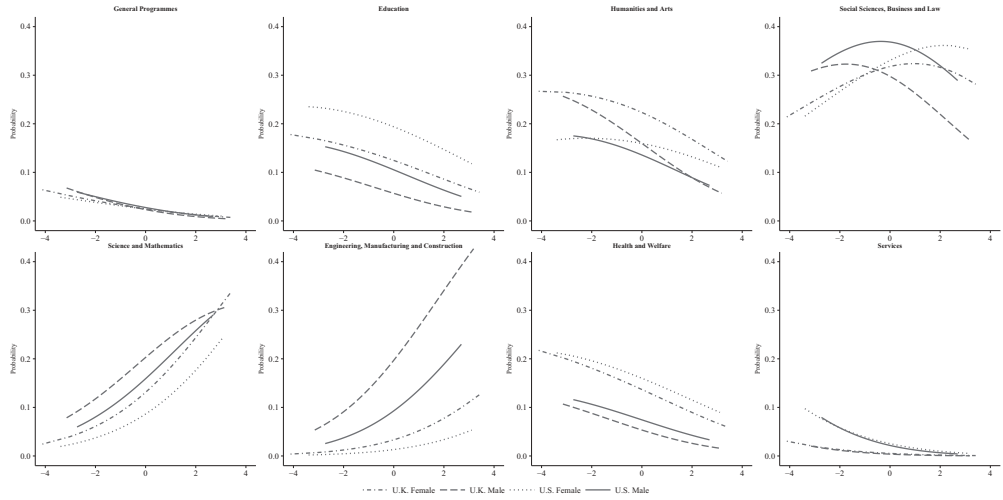
Specifically, the average marginal effect of numeracy (literacy) index is associated with one-standard-deviation increase in numeracy (literacy) skill index. The reference field of study is humanities, language, and arts.

business, and law becomes positive with respect to the increase in either skill. Therefore, the observed non-monotonicity may arise from the variations of intra-field choices as skills increase.

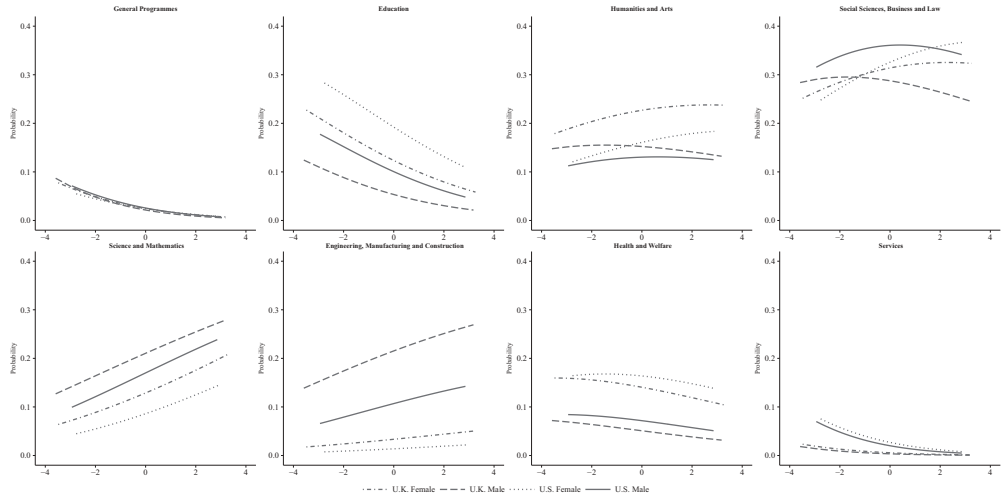
Men and women in both countries exhibit quite congruent change in choices of major against either numeracy skills or literacy skills, though different in the level. By that I mean, relation between skills and field choices is very similar between US and UK. On the other hand, this relation differs substantially when comparing male and female. Data suggest that this relation is more gender-specific and is stable across countries.¹⁶

However, literacy serves as a weaker predictor of field choices in STEM majors. For instance, the average increase in possibility of choosing engineering, manufacturing, and construction is as three times associated with one-standard-deviation increase in numeracy as literacy, as shown in Table 4 Column 3 and 5, where

¹⁶We leave one exception as in the field of services. It is likely to be due to its precision as its sample size is very small.



(a) Numeracy Index



(b) Literacy Index

Figure 2. Field choice prediction by skill index.

I estimate the average marginal effect of one-standard-deviation increase in numeracy and literacy skill on field choices, respectively.

3.2. Field premium estimates

I first adopt Equation (5) to estimate the field premiums,

$$\ln W_i = \text{Constant} + \sum_{j=1}^{J-1} \alpha_j \cdot 1\{J_i = j\} + \varphi_C \cdot \mathbf{C}_i + \epsilon_i \quad (5)$$

where W_i is the observed monthly earnings, PPP adjusted, of individual i , $1\{J_i = j\}$ is the binary indicator of choosing field j , which takes value 1 if $J_i = j$ and 0 otherwise. Let C_i be a vector of control variables, including country origin, gender, age and a quadratic polynomial of work experience (Mincer, 1958).¹⁷ The idiosyncratic error term, ϵ_i , is assumed to be statistically independent of J_i and C_i . Equation (6) extends Equation (5) by accounting for unevenly distributed skills across fields. It includes basic cognitive skills, denoted as S_i .¹⁸ The model now reads as follows,

$$\ln W_i = \text{Constant} + \sum_{j=1}^{\mathcal{J}-1} \beta_j \cdot 1\{J_i = j\} + \varphi_S \cdot S_i + \varphi_C \cdot C_i + \epsilon_i \quad (6)$$

The standard errors in both models are clustered by the fields of study, gender and country. The parameters, α_j and β_j ($j = 1, 2, \dots, \mathcal{J}$), in Equations (5) and (6) are of interest. Coefficient α_j measures the observed earnings premium for field j while β_j measures the adjusted premium when we correct for skills. The difference $\alpha_j - \beta_j$ indicates the extent to which the difference between cognitive skill distributions across fields of study can influence earnings. In all my models, the field of humanities and arts is set to be the reference group.¹⁹ The coefficient of each field of study represents the fixed effect compared to humanities and arts.

Model 1 in the Table 5 Column 2 reports estimates of Equation (5). The results are in line with our expectation of the earnings gap across fields of study. Health and welfare has a premium of 30% and science and mathematics 23%. Surprisingly, engineering, manufacturing, and construction only yields 18.4% premium compared to the reference group. Field of humanities and arts yields the lowest.

In Column 3, the numeracy index is added to the model as a proxy of numeracy skills. Since each skill measurement is standardized to (0,1), the coefficient can, therefore, be interpreted as 12.9% increase in monthly earnings associated with one-standard-deviation increase in numeracy skills.

A comparison of results in Column 2 and 3 shows that the field premiums change after conditional on numeracy skills. Engineering, manufacturing, and construction has the highest reduction in the field premium with 35.7% to $\beta_j = 0.119$. Science and mathematics also has a large reduction with 23% to $\beta_j = 0.177$. Numeracy skills explain up to two-fifths of field premiums for STEM majors. However, field of services demonstrates around one-third increase in premium. After controlling for numeracy skills, large premiums do exist, for example, for the field of health and welfare. Figure 3 demonstrates this point more clearly, which presents the percentage change in field premiums after adjusting for numeracy index.

I add literacy index as my new control in Column 4. Literacy has around one-quarter smaller return than numeracy and it almost has no effect of explaining STEM majors'

¹⁷The US distinguishes between undergraduate and graduate degree clearly but in the UK, undergraduate and graduate degrees are also aggregated. Since graduate degree indicator would measure different types of schooling in the US and the UK, adding it may incur systematic error. Also, since some majors may open the door to more lucrative graduate school options, such as MBA, than others, the return to graduate degree should be considered as an element of the returns to major instead of as controls (Altonji et al., 2014). Therefore, I do not control for graduate degree in my model.

¹⁸I use numeracy and literacy skill indices in my model. For simplification, I use S_i^L and S_i^N to refer to individual i 's literacy and numeracy skill indices, respectively. That is, $S_i = [S_i^L, S_i^N]$.

¹⁹In the rest of my paper, when I use "field premium" or "return to certain field", I refer to the fixed effect of that field compared to that of humanities and arts.

Table 5. Field premium estimates by field of study in term of log monthly earnings, PPP adjusted.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
<i>STEM Majors</i>	2	3	4	5	6	7
Science and Mathematics	0.230*** (0.06)	0.177** (0.06)	0.217*** (0.06)	0.176** (0.06)	0.176** (0.06)	0.176** (0.06)
Engineering, Manufacturing and Construction	0.184** (0.06)	0.119+ (0.06)	0.173** (0.06)	0.118+ (0.06)	0.118+ (0.06)	0.118+ (0.06)
<i>Non-STEM Majors</i>						
General Programs	0.146 (0.09)	0.168+ (0.10)	0.184* (0.09)	0.167+ (0.10)	0.167+ (0.10)	0.167+ (0.10)
Education	0.028 (0.07)	0.033 (0.07)	0.051 (0.07)	0.033 (0.07)	0.033 (0.07)	0.033 (0.07)
Social Science, Business and Law	0.252*** (0.06)	0.236*** (0.06)	0.252*** (0.06)	0.235*** (0.06)	0.235*** (0.06)	0.235*** (0.06)
Health and Welfare	0.300*** (0.07)	0.308*** (0.07)	0.310*** (0.07)	0.308*** (0.07)	0.308*** (0.07)	0.308*** (0.07)
Services	0.165 (0.12)	0.212+ (0.11)	0.211+ (0.12)	0.212+ (0.12)	0.212+ (0.12)	0.212+ (0.12)
<i>Returns to Skill</i>						
Numeracy Index		0.129*** (0.02)		0.132*** (0.03)		0.129*** (0.02)
Literacy Index			0.098*** (0.02)	-0.003 (0.03)	0.101*** (0.02)	
$\text{Proj}_{S^L}^\perp(S^N)$					0.081*** (0.02)	
$\text{Proj}_{S^N}^\perp(S^L)$						-0.002 (0.02)
R^2	0.171	0.193	0.184	0.193	0.193	0.193

+ $p < .1$, * $p < .05$, ** $p < .01$, *** $p < .001$

This table estimates the field premiums in term of log monthly earnings, PPP adjusted. The data sample consists of $N = 2469$, with both US and UK data samples. Individuals in agriculture and veterinary major are excluded due to limited number of observations. All regressions in this table controls for country, gender, age and a quadratic in work experience. Field of humanities and arts is set to be my base. The work experience and squared work experience are also standardized to zero-mean and one-standard-deviation by country. Both numeracy and literacy skills are standardized by country and age category. The numeracy index refers to the standardized numeracy skill scores measured by PIAAC and the literacy index refers to the standardized literacy skill scores. $\text{Proj}_{S^L}^\perp(S^N)$ refers to the standardized component of numeracy index orthogonal to literacy index; $\text{Proj}_{S^N}^\perp(S^L)$ refers to the standardized component of literacy index orthogonal to numeracy index. Standard errors are clustered by country, gender, age, and field of study and are presented in the brackets.

premiums. Column 5 includes both literacy and numeracy skills and the return to literacy is negligibly negative compared to numeracy. Taking high correlation between two skill indices ($\rho = 0.79$) into consideration, I attempt to further evaluate the relative importance of these skill components in the labor market.

Let S_i^L and S_i^N denote literacy and numeracy skills for individual i respectively. Let $S^L = [S_i^L]_{i \in \mathcal{I}}$ be the vector of literacy skills, generated by stacking literacy index, and let $S^N = [S_i^N]_{i \in \mathcal{I}}$ be the vector of numeracy skills. I decompose numeracy into a component along literacy and a component orthogonal to literacy. The explicit formulas are given by

$$\text{Proj}_{S^L}(S^N) = \frac{\langle S^N, S^L \rangle}{\langle S^L, S^L \rangle} \cdot S^L \quad (7)$$

$$\text{Proj}_{S^L}^\perp(S^N) = S^N - \text{Proj}_{S^L}(S^N) \quad (8)$$

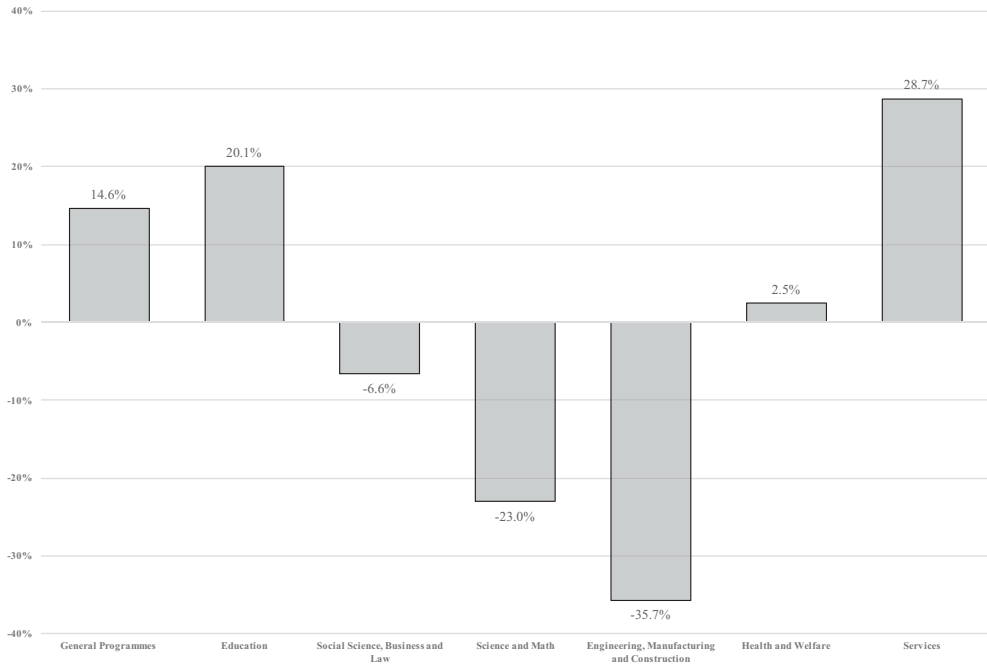


Figure 3. Change in field premium, adjusted by numeracy index.

where $\langle \mathbf{x}, \mathbf{y} \rangle$ is the Euclidean inner product between \mathbf{x} and \mathbf{y} . Similarly, literacy is decomposed with respect to numeracy.

$$\text{Proj}_{\mathbf{S}^N}(\mathbf{S}^L) = \frac{\langle \mathbf{S}^L, \mathbf{S}^N \rangle}{\langle \mathbf{S}^N, \mathbf{S}^N \rangle} \cdot \mathbf{S}^N \quad (9)$$

$$\text{Proj}_{\mathbf{S}^{N\perp}}(\mathbf{S}^L) = \mathbf{S}^L - \text{Proj}_{\mathbf{S}^N}(\mathbf{S}^L) \quad (10)$$

I include literacy skills and the component of numeracy orthogonal to literacy in Column 6. The return to the latter is 0.081. This amount is comparable to the return to literacy ($\varphi_{SL} = 0.101$). On the other hand, I use numeracy skills and the orthogonal component of literacy with respect to numeracy in Column 7. The return to the latter is negligibly negative. Even though both skills may share the same latent cognitive trait, numeracy is of more importance than literacy in determining one's labor market outcomes.

Overall, my results do show that a sizable portion of the field premiums is due to heterogeneous distribution of cognitive skills across fields. The difference drives different choices of major and consequently after controlling for skills, the return to each field of study is altered. The effects of cognitive skills are not homogeneous in all dimensions. Considering skills as factors of production, their productivities are estimated through φ_{SN} for numeracy and φ_{SL} for literacy. One-standard-deviation increase in numeracy is associated with more productivity gain than in literacy. Field premiums are estimated by field fixed effects and the field of health and welfare presents large premiums among all fields.

3.3. Robustness checks

In this section, I will address two concerns in my baseline results. First, I test if my results would be biased by the sensitivity in earnings on the career trajectory. Second, I assess the influence of the income data attrition (up to 7% missing) on my results.

The entry-age workers may be statistically discriminated (Altonji & Pierret, 2001). They argue that with little information of workers' productivity at the early stage of their careers, firms can only distinguish workers based on their easily observed characteristics such as education. Firms would also be likely to decrease their wages to deal with the uncertainty in workers' quality. Therefore, entry-age workers' wages may be underrepresented, incurring measurement errors in estimating field premiums as well as the returns to skills. On the other hand, earning function might be different for workers who are prepared for the retirement. Including these people may introduce model specification error.

One possible solution is to exclude these workers from my data and restrict to prime-age workers. The prime-age workers are defined as those who are between 35 and 54 years old and their current earnings can be treated as consistent proxy to the lifetime earnings (Böhlmark & Lindquist, 2006; Haider & Solon, 2006; Hanushek, Schwerdt, Wiederhold, & Woessmann, 2015). The results are shown in Table 6 Column 2–3 and the sample reselection does not impair my baseline results.

Second, the attrition of income data is small but it corresponds mostly to nonresponse. The decision to response or not is affected by other unobserved characteristics. People who are in a lower social position, with a lower educational attainment, female, African American, or older in age have higher item nonresponse rate (Lor, Bowers, Krupp, & Jacobson, 2017; Ross & Reynolds, 1996). Beatty and Herrmann (2002) suggest that other cognitive and motivational factors also contribute to the income nonresponse, such as cognitive ability to remember and retrieve relevant information to report, judgment of the level of precision required by the survey and one's willingness to answer. The motivation can be affected by one's perception of "social distance" and mistrust (Johnson, O'Rourke, Burris, & Owens, 2002; Ross & Reynolds, 1996). Thus, the nonresponse might create potential for bias in estimation, as it is a nonrandom selection (Dillman, Eltinge, Groves, & Little, 2002).

To correct for the possible bias, I use Inverse Probability Weighting method to estimate the fixed effects of each field.²⁰ The results are shown in Table 6 Column 4–7. I use Model 1 and Model 2 in Table 5 for comparison. In Columns 4 and 5, I use the inverse of the propensity score to weight the data, the matching method proposed by Rosenbaum and Rubin (1983) to reduce bias. I also follow Hirano, Imbens, and Ridder (2003) and weight by the inverse of a nonparametric estimate of the propensity score to obtain more efficient estimates. The estimates are in Columns 6 and 7. Using both IPW methods does not affect my baseline results. The observed absence of nonresponse bias may partly due to my sample selection, as I focus on the native respondents with college degree or above. These people are less likely to be in the higher-nonresponse-rate group. Therefore, the missing in income data can be treated as random and nonresponse bias is not of major concerns.

²⁰Literature that use IPW method to deal with missing data include not limited to Prokhorov and Schmidt (2009); Tsiatis (2006); Wooldridge (2007).

Table 6. Robustness checks for Model 1 and 2 in Table 5.

	Prime-Age Respondents		Weighted acc. Matching Approach		Weighted acc. Hirano et al. (2003)	
	2	3	4	5	6	7
<i>STEM Majors</i>						
Science and Mathematics	0.296*** (0.07)	0.249*** (0.07)	0.224*** (0.06)	0.166** (0.06)	0.223*** (0.06)	0.166* (0.06)
Engineering, Manufacturing and Construction	0.221** (0.07)	0.140* (0.07)	0.162* (0.06)	0.092 (0.07)	0.166** (0.06)	0.102 (0.06)
<i>Non-STEM Majors</i>						
General Programs	0.100 (0.13)	0.148 (0.14)	0.147 (0.10)	0.169 ⁺ (0.10)	0.139 (0.10)	0.156 (0.10)
Education	0.125 ⁺ (0.07)	0.135 ⁺ (0.07)	0.015 (0.07)	0.017 (0.07)	0.020 (0.07)	0.023 (0.07)
Social Science, Business and Law	0.337*** (0.06)	0.318*** (0.06)	0.242*** (0.06)	0.225*** (0.06)	0.238*** (0.06)	0.222*** (0.06)
Health and Welfare	0.332*** (0.07)	0.348*** (0.08)	0.270*** (0.08)	0.278*** (0.08)	0.296*** (0.07)	0.299*** (0.07)
Services	0.211 (0.15)	0.240 (0.15)	0.158 (0.12)	0.203 ⁺ (0.11)	0.154 (0.12)	0.193 ⁺ (0.12)
<i>Returns to Skill</i>						
Numeracy Index		0.134*** (0.02)		0.136*** (0.02)		0.128*** (0.02)
R^2	0.180	0.204	0.166	0.189	0.166	0.187
No. Obs.	1248	1248	2467	2467	2467	2467

+ $p < .1$, * $p < .05$, ** $p < .01$, *** $p < .001$

This table evaluates Models 1 and 2 from Columns 2 and 3 in Table 5. All regressions in this table controls for country, gender, age, and a quadratic in work experience. The work experience and squared work experience are also standardized to zero-mean and one-standard-deviation by country. Standard errors are clustered by country, gender, age, and field of study and are presented in the brackets. Column 2 and 3 presents estimates for prime-age respondents (Aged from 35 to 54). Columns 4–7 present IPW estimates. To estimate weight, I use a model whose dependent variable is indicator of non-missing in income ($1\{\text{individual } i \text{ reports } W_i\}$) and whose predictors are factors that may contribute to income nonresponse (Beatty & Herrmann, 2002; Johnson et al., 2002; Ross & Reynolds, 1996). I estimate the weight for the US and UK separately. For the US, I include gender indicator, education year, number of household members, age, race/ethnicity, region, standardized numeracy skill index, standardized factor score on learning strategy, and standardized self-reported score on if trust others to estimate the corresponding weight. For the UK, I include gender indicator, education year, number of household members, age, standardized numeracy skill index, standardized factor score on learning strategy, and standardized self-reported score on if trust others to estimate the corresponding weight. In Columns 4 and 5, I use the inverse of the weight estimated from a logit model in which all covariates are standardized, the method described by Rosenbaum and Rubin (1983). In Columns 6 and 7, I follow Hirano et al. (2003) and weight by the inverse of a nonparametric estimate of the propensity score.

4. Discussion and conclusion

The earnings gap across college majors is substantial and is still being widened. This paper attempts to shed some light on the understanding of how field premiums and earnings gap across majors are generated. Benefiting from the rich data from PIAAC, I mainly focus on numeracy and literacy skills and examine their roles in the earnings gap across fields of study and I have the following results.

First of all, heterogeneity in cognitive skills affects people's college field choices. The increase in either numeracy or literacy shifts people away from fields such as education, humanities and arts toward high-skill subjects, such as STEM majors. The one exception is for women regarding humanities and arts. The increase in numeracy skills shifts women away from humanities and arts while the increase in literacy skill shifts them toward it. This striking similarity in effects of both skills strongly suggests that numeracy and literacy simply present different facets of the same latent cognitive trait. For both skills, I find that

the possibility of choosing certain field with increase in either skill is more gender-specific than country-specific, except in the field of services. Namely, I find much greater tendency for men than for women to STEM majors regardless of skill level. Particularly, in the field of engineering, the striking difference in the marginal effects of increase in skills – illustrated in [Figure 2](#) as the slope of the curve of predicted possibility of any field choice in response to skill change – reveals how differently men and women take skill into account when choosing engineering. It might suggest that men have a relatively stronger preference for that field over women.²¹ We also observe a non-monotonic relationship between choosing the field of social science, business, and law and both skills. One of the issues may be the intra-field skill difference resulting from the complexity in the field classification by ISCED, but other reasons may also lie behind.

Secondly, the return to human capital acquired through educational process within one particular field should be considered as the actual field premium. Different college majors provide students with different course contents and emphasize on different practical skills and knowledge. The difference in college course contents further drive variation in occupational choices in the labor market. Nevertheless, that field premium is masqueraded by the self-selection due to heterogeneous skill distribution. After controlling for numeracy skills, STEM premiums faces up to two-fifths reduction. In other words, the returns to cognitive skills explain a sizable portion of STEM premiums and educational process in STEM does not provide as many monetary returns as observed. On the other hand, most non-STEM majors, except the field of social science, business, and law, increase the relative premiums.

However, large field premiums do exist even after skill heterogeneity is corrected. It suggests that premium is not merely generated via sorting people with high cognitive skills into more lucrative fields and awarding for their high skills. To investigate why, it is unlikely that the high premium is due to a third type skill orthogonal to both literacy and numeracy that I do not have access to. One possible reason lies in the likelihood of mismatch between employment and field of study in college, i.e., how one's job is related to the field of study. Since occupational mismatched workers earn lower wages than well-matched workers with the same degree field, the large premium differentiation can be explained by the extent to which workers in each field are mismatched (Robst, 2007). People in certain fields such as health professions, engineering, business, and law are more likely to be in a field-related job and low mismatch prevalence rate in these fields might explain their relatively high premiums (Altonji et al., 2012; Grubb, 1997; Robst, 2007). It also might be the case that skills learned in those fields are highly valued by the market. For instance, the field premium of health and welfare even overshadows the premium for engineering both with and without conditioning on numeracy skills, the result consistent with Kirkeboen, Leuven, and Mogstad (2016), who show that the payoff to medicine is the largest. We could regard its premium as occupational rewards available in the labor market for its practical skills and knowledge learned in college. On the other hand, the observed rewards may be also due to regulation of healthcare market (Schoen et al., 2010).

Furthermore, numeracy is found to be an important ability for labor market outcomes than literacy. The finding of relative importance of numeracy skills over literacy skills

²¹My result complies with complies with Turner and Bowen (1999).

agrees with Arcidiacono (2004) and Paglin and Rufolo (1990), who use SAT and/or GRE scores as proxies to verbal and mathematics abilities. Numeracy is of a scarce attribute and its return is observed to be increasing over year (Grogger & Eide, 1995; Paglin & Rufolo, 1990). To be noted, numeracy skills include not only quantitative ability but also literacy components that correspond to “access, use, interpret and communicate mathematical information and ideas” (OECD, 2013). My results suggest that numeracy skill is a better predictor of earnings than mathematics ability or literacy skills alone.

The analysis in this paper has some limitations. First, my model on field choices does not take other residual forces into account, such as preferences and tastes, expected lifetime earnings which are also play a role in one’s choice of major (Altonji et al., 2012; Arcidiacono et al., 2012; Berger, 1988). Since respondents were already in the labor market when they completed PIAAC survey, it is unlikely that PIAAC would have questions on what their college field preferences were several years ago. Second, my measurement of labor market outcomes mainly considers the pecuniary returns but does not include the non-pecuniary benefits. The latter is of more importance to women (Zafar, 2013). For instance, one might put more value on familial responsibility or on satisfaction from cooperating with other people than monthly earnings (Diekmann, Brown, Johnston, & Clark, 2010; Turner & Bowen, 1999). However, due to the limitations of PIAAC questionnaires, it is challenging to examine these non-pecuniary benefits. Third, even though PIAAC provided monetary incentives to increase response rates and attempted to increase the accuracy of cognitive measurement by using the population model, the scores may still partly reflect non-cognitive components, such as efforts, attitudes, and personal traits (Heckman & Kautz, 2014; OECD, 2013; PIAAC Numeracy Expert Group, 2009). It would be unavoidable in almost all IQ tests.

This paper attempts to answer the question of how to understand the variations in choices of college major by taking uneven distribution of cognitive skills into account. It shows that a well-measured cognitive index can explain much of one’s choices of major in college as well as the current earnings gap across fields of study.

Acknowledgments

I am specially grateful to Rodrigo Pinto for his helpful comments throughout all stages of writing. All errors are my own.

Disclosure statement

No potential conflict of interest was reported by the author.

Notes on contributor

Kan Yao is a student in Economics at the University of California, Los Angeles, at the time this project is completed. His fields of interest include labor economics and econometrics.

References

- Allen, J., & Van der Velden, R. (2001). Educational mismatches versus skill mismatches: Effects on wages, job satisfaction, and on-the-job search. *Oxford Economic Papers*, 53(3), 434–452.
- Altonji, J. G., Blom, E., & Meghir, C. (2012, April). *Heterogeneity in human capital investments: High school curriculum, college major, and careers* (Working Paper 17985). National Bureau of Economic Research.
- Altonji, J. G., Kahn, L. B., & Speer, J. D. (2014, May). Trends in earnings differentials across college majors and the changing task composition of jobs. *American Economic Review: Papers & Proceedings*, 104(5), 387–393.
- Altonji, J. G., & Pierret, C. R. (2001). Employer learning and statistical discrimination. *The Quarterly Journal of Economics*, 116(1), 313–350.
- Arcidiacono, P. (2004). Ability sorting and the returns to college major. *Journal of Econometrics*, 121(1–2), 343–375.
- Arcidiacono, P., Hotz, V. J., & Kang, S. (2012). Modeling college major choices using elicited measures of expectations and counterfactuals. *Journal of Econometrics*, 166(1), 3–16.
- Beatty, P., & Herrmann, D. (2002). To answer or not to answer: Decision processes related to survey item response. In D. A. Dillman, J. L. Eltinge, R. M. Groves, & R. J. Little Eds., *Survey nonresponse*, Chapter 5 (pp. 71–85). New York: John Wiley & Sons, Inc.
- Berger, M. C. (1988). Predicted future earnings and choice of college major. *Industrial and Labor Relations Review*, 41(3), 418–429.
- Böhlmark, A., & Lindquist, M. J. (2006). Life-cycle variations in the association between current and lifetime income: Replication and extension for Sweden. *Journal of Labor Economics*, 24(4), 879–896.
- Carnevale, A. P., Cheah, B., & Hanson, A. R. (2015). *The economic value of college majors*. Washington: Georgetown University Center on Education and the Workforce.
- Cunha, F., & Heckman, J. J. (2009, April–May). The economics and psychology of inequality and human development. *Journal of the European Economic Association*, 7(2–3), 320–364.
- Diekmann, A. B., Brown, E. R., Johnston, A. M., & Clark, E. K. (2010). Seeking congruity between goals and roles: A new look at why women opt out of science, technology, engineering, and mathematics careers. *Psychological Science*, 21(8), 1051–1057.
- Dillman, D. A., Eltinge, J. L., Groves, R. M., & Little, R. J. (2002). Survey nonresponse in design, data collection, and analysis. In D. A. Dillman, J. L. Eltinge, R. M. Groves, & R. J. Little Eds., *Survey nonresponse*, Chapter 1 (pp. 3–26). New York: John Wiley & Sons, Inc.
- Duckworth, A. L., Quinn, P. D., & Tsukayama, E. (2012). What no child left behind leaves behind: The roles of IQ and self-control in predicting standardized achievement test scores and report card grades. *Journal of Educational Psychology*, 104(2), 439–451.
- Gal, I., & Tout, D. (2014, May). *Comparison of PIAAC and PISA frameworks for numeracy and mathematical literacy* (OECD Education Working Papers 102). OECD Publishing.
- Grogger, J., & Eide, E. (1995). Changes in college skills and the rise in the college wage premium. *The Journal of Human Resources*, 30(2), 280–310.
- Grubb, W. N. (1997). The returns to education in the sub-baccalaureate labor market, 1984–1990. *Economics of Education Review*, 16(3), 231–245. Education and work, and efficiency in education: Essays in memory of Charles Scott Benson.
- Haider, S., & Solon, G. (2006). Life-cycle variation in the association between current and lifetime earnings. *American Economic Review*, 96(4), 1308–1320.
- Hanushek, E. A., Schwerdt, G., Wiederhold, S., & Woessmann, L. (2015). Returns to skills around the world: Evidence from PIAAC. *European Economic Review*, 73(C), 103–130.
- Heckman, J., Carneiro, P., & Cunha, F. (2004). *The technology of skill formation* (2004 Meeting Papers 681, Society for Economic Dynamics).
- Heckman, J. J., & Kautz, T. (2014). Fostering and measuring skills: Interventions that improve character and cognition. In J. J. Heckman, J. E. Humphries, & T. Kautz Eds., *The myth of achievement tests: The GED and the role of character in American life*, Chapter 9 (pp. 341–430). Chicago: University of Chicago Press.

- Heckman, J. J., Stixrud, J., & Urzúa, S. (2006, July). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, 24(3), 411–482.
- Hirano, K., Imbens, G. W., & Ridder, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica*, 71(4), 1161–1189.
- Hopkins, K. D., & Bracht, G. H. (1975). Ten-year stability of verbal and nonverbal IQ scores. *American Educational Research Journal*, 12(4), 469–477.
- Johnson, T. P., O'Rourke, D., Burris, J., & Owens, L. (2002). Culture and survey nonresponse. In D. A. Dillman, J. L. Eltinge, R. M. Groves, & R. J. Little Eds., *Survey nonresponse*, Chapter 4 (pp. 55–69). New York: John Wiley & Sons, Inc.
- Kirkeboen, L. J., Leuven, E., & Mogstad, M. (2016). Field of study, earnings, and self-selection*. *The Quarterly Journal of Economics*, 131(3), 1057–1111.
- Lor, M., Bowers, B., Krupp, A., & Jacobson, N. (2017). Tailored explanation: A strategy to minimize nonresponse in demographic items among low-income racial and ethnic minorities. *Survey Practice*, 10(3), 1–11.
- McFadden, D. (1973). Conditional logit analysis of qualitative choice behaviour. In P. Zarembka (Ed.), *Frontiers in econometrics* (pp. 105–142). New York, NY, USA: Academic Press New York.
- Mincer, J. (1958). Investment in human capital and personal income distribution. *Journal of Political Economy*, 66(4), 281–302.
- Mincer, J. (1974). *Schooling, experience, and earnings*. Cambridge: National Bureau of Economic Research, Inc.
- OECD (2009). Sample questions and questionnaire. Retrieved from <https://www.oecd.org/skills/piaac/samplequestionsandquestionnaire.htm>
- OECD (2013). *Technical report of the Survey of Adult Skills (PIAAC)*. Paris, France: OECD.
- Paglin, M., & Rufolo, A. M. (1990). Heterogeneous human capital, occupational choice, and male-female earnings differences. *Journal of Labor Economics*, 8(1), 123–144.
- PIAAC Numeracy Expert Group. (2009). *PIAAC numeracy: A conceptual framework* (Number 35 in OECD Education Working Papers). OECD Publishing.
- Prokhorov, A., & Schmidt, P. (2009). Gmm redundancy results for general missing data problems. *Journal of Econometrics*, 151(1), 47–55.
- Robst, J. (2007). Education and job match: The relatedness of college major and work. *Economics of Education Review*, 26(4), 397–407.
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41–55.
- Ross, C. E., & Reynolds, J. R. (1996). The effects of power, knowledge, and trust on income disclosure in surveys. *Social Science Quarterly*, 77(4), 899–911.
- Schoen, C., Osborn, R., Squires, D., Doty, M. M., Pierson, R., & Applebaum, S. (2010). How health insurance design affects access to care and costs, by income, in eleven countries. *Health Affairs*, 29(12), 2323–2334.
- Tsiatis, A. (2006). Semiparametric theory and missing data. In *Springer series in statistics*. New York: Springer-Verlag New York.
- Turner, S., & Bowen, W. G. (1999). Choice of major: The changing (unchanging) gender gap. *ILR Review*, 52(2), 289–313.
- UNESCO. (2012). *International standard classification of education ISCED 2011*. Quebec, Canada: UNESCO Institute for Statistics.
- van de Werfhorst, H. G. (2002). Fields of study, acquired skills and the wage benefit from a matching job. *Acta Sociologica*, 45(4), 287–303.
- Wooldridge, J. M. (2007). Inverse probability weighted estimation for general missing data problems. *Journal of Econometrics*, 141(2), 1281–1301.
- Zafar, B. (2013). College major choice and the gender gap. *The Journal of Human Resources*, 48(3), 545–595.

Appendix A. Descriptive statistics

A.1. Data Sample

The breakdown by country, employment status, and income data availability are presented in Figure A1. The sample consists of prime-age native respondents with a tertiary degree or above who are active in the labor market.

A.2. Skills

The distribution of literacy and numeracy skill score for each field of study within my data sample ($N = 2485$) are presented in Figure A2. The US and UK data are combined. The x-axis presents the literacy/numeracy skill raw scores each respondent can get and a more detailed interpretation of the raw scores is in PIAAC Technical Report Section 5 Chapter 21 (OECD, 2013). The average literacy/numeracy skill score for the whole data sample is presented as the reference line in each plot.

Appendix B. Classification of Field Education and Training

According to <http://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-isced-2011-en.pdf> International Standard Classification of Education ISCED 2011, This section presents detailed classification of field education and training (UNESCO, 2012). Majors are organized in 9 broad groups.

- (1) General programs
It includes basic general programs pre-primary, elementary, primary, secondary and etc; simple and functional literacy or numeracy training; personal development such as life orientation programs.
- (2) Teacher training and education science
Teaching training includes those for different levels of education, such as pre-school, kindergarten, elementary school, vocational, etc.;

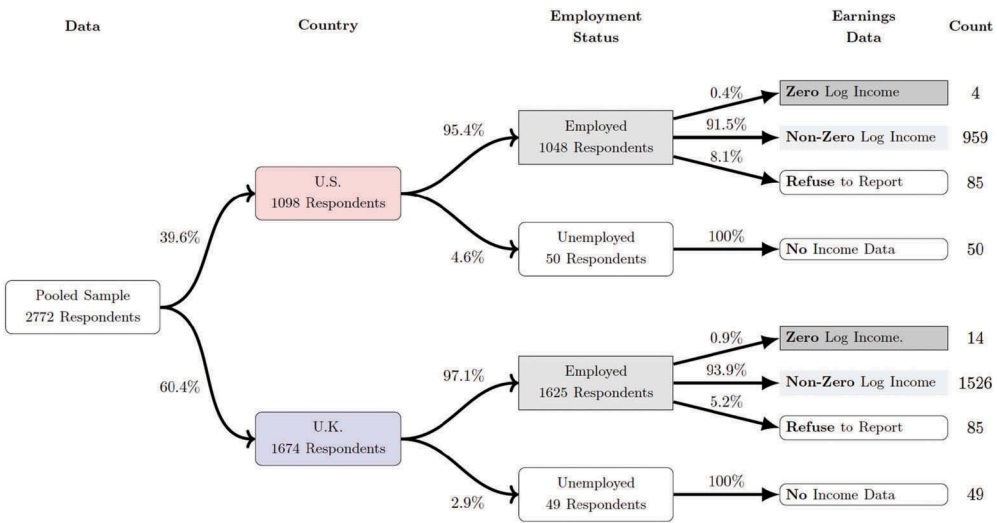
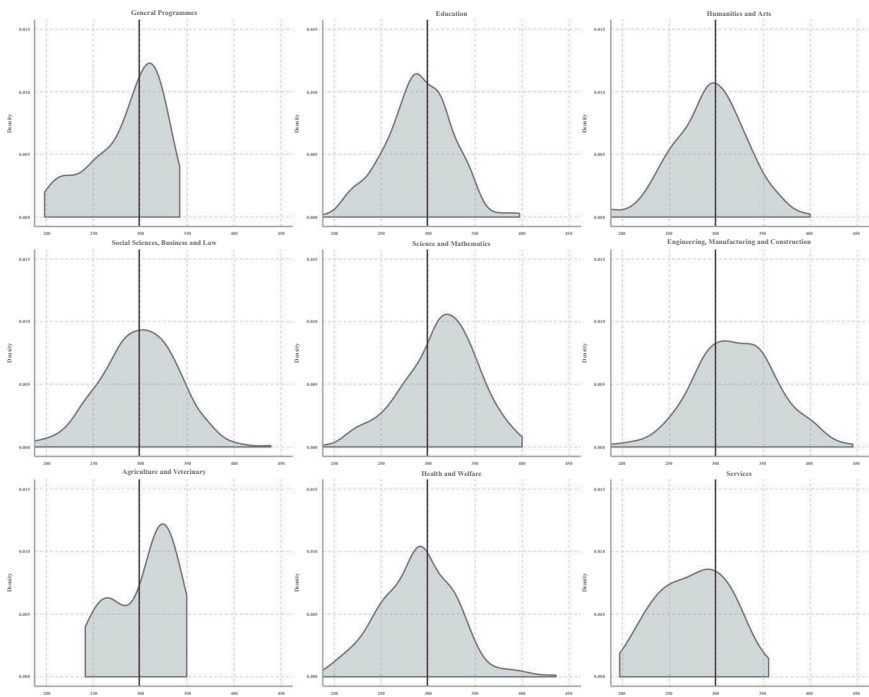
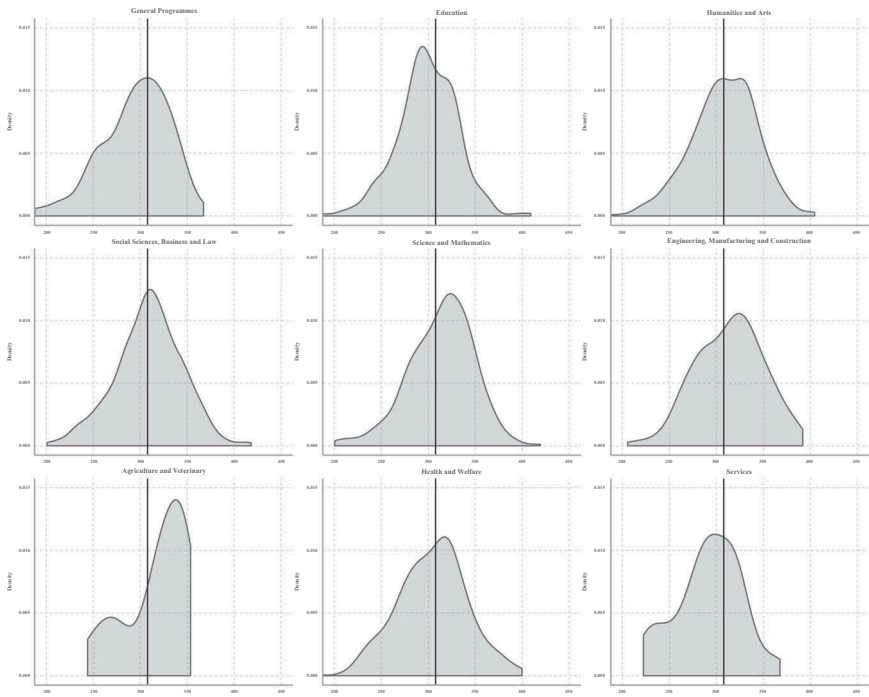


Figure A1. Breakdown of respondents by country, employment status, and income data availability.



(a) Literacy



(b) Numeracy

Figure A2. Skill score distribution by field of study.

Education science includes curriculum development, educational assessment, testing and measurement, etc.

(3) Humanities and arts

Humanities include religion and theology, native and foreign languages and cultures and other subjects such as linguistics, comparative literature, history, archaeology, philosophy, ethics.

Arts include fine arts, performing arts, graphic and audio-visual arts, design and craft skills.

(4) Social sciences, business, and law

It contains four sub-fields: social and behavioral science, journalism and information, business and administration and law.

(5) Science

It has four sub-fields: life sciences, physical sciences, mathematics and statistics, and computer science.

(6) Engineering, manufacturing, and construction

It contains engineering and engineering trades, manufacturing and processing, and architecture and building. For instance, manufacturing and processing includes food and drink processing, textiles, clothes, footwear, leather, materials (wood, paper, plastic, glass, etc.), mining, and extraction. Civil engineering belongs to the last sub-field.

(7) Agriculture and veterinary

It includes agriculture, forestry and fishery and veterinary, such as animal husbandry, horticulture and gardening, forestry and forest product techniques, veterinary medicine, etc.

(8) Health and welfare

Health field includes medicine, medical services, nursing, and dental services; welfare and social services contains social care, such as care of the disabled, childcare, youth services, gerontological services, and social work.

(9) Services

It includes four sub-fields: personal services, transport services, environmental protection and security services. For instance, personal services are hotel and catering, travel and tourism, sports and leisure, hairdressing, beauty treatment, etc.