THREE ESSAYS ON HUMAN CAPITAL: DEVELOPMENT AND IMPACT

A Dissertation
Presented to
The Academic Faculty

by

Tingting Tong

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Economics

Georgia Institute of Technology
August 2016

THREE ESSAYS ON HUMAN CAPITAL: DEVELOPMENT AND IMPACT

Approved by:

Dr. Haizheng Li, Advisor
School of Economics
Georgia Institute of Technology

Dr. Usha Nair-Reichert
School of Economics
Georgia Institute of Technology

Dr. Patrick S. McCarthy
School of Economics
Georgia Institute of Technology

Dr. Olga Shemyakina
School of Economics
Georgia Institute of Technology

Dr. Samuel Greiff
Educational Assessment
University of Luxembourg

Date Approved: April 29th, 2016

To my family and my advisor

ACKNOWLEDGEMENTS

My deepest gratitude goes to my advisor, Professor Haizheng Li, for his excellent guidance, continuous support, and patience. I am extremely fortunate to have him as my advisor. He always has time for me despite his own schedule. He worked closely with me at daily basis and helped me solve numerous issues in my dissertation. Without his guidance, this dissertation would not have been finished. I am also grateful to him for offering me valuable suggestions when I was in the job market. His serious research attitudes and self-discipline set an excellent example for me in my future career.

My sincere thanks also goes to Professor Usha Nair-Reichert, Professor Patrick S. McCarthy, Professor Olga Shemyakina, and Professor Samuel Greiff for serving as my committee members. Professor Usha Nair-Reichert gave me a lot of valuable suggestions, especially when I was trying to set up a theoretical model for my leadership study. I am also thankful for the insightful comments from Professor Patrick S. McCarthy, which gave me a deeper understanding of leadership literature and the contribution of my study. Many thanks to Professor Olga Shemyakina, whose comments helped me improve my dissertation. I am also grateful for the comments and editions from Professor Samuel Greiff, especially on the leadership paper.

I am also indebted to Professor Vivek Ghosal, Professor David Laband, Professor Michael Kummer, and Professor Shatakshee Dhongde for their valuable comments during my proposal presentation and afterwards. Last but not least, I would like to thank my parents, my grandparents, my sister, and my boyfriend for their spiritual support throughout my Ph.D study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

This dissertation investigates the development and impact of human capital, which includes three essays. The first essay relates to the impact of human capital. We conduct an economic analysis about the impact of an individual's human capital on the potential of becoming a leader. The data come from the recent Programme for the International Assessment of Adult Competencies survey (PIAAC). Our human capital indicators include not only traditional measures such as education and experience, but also various measures of cognitive and noncognitive ability. We specifically investigate the effect of imperfect ability measurement and possible reverse causality on the estimation results. We find that, among various cognitive and noncognitive ability measures, problem-solving ability and perseverance are the most important in affecting an individual's potential leadership.

The second essay relates to the Massive Open Online Courses (MOOC), which has become an important trend gathering increasing attentions of higher education since 2012. In this essay, we evaluates factors affect the demand of MOOC in OECD countries and in China from 2012 to 2015 with the application of big data. We use Google Trends and Baidu Index to proxy MOOC demand. Based on the classic demand theory, factors affecting MOOC demand we consider include tuition, internet speed, unemployment rate, education level, income, and population. In both studies, higher unemployment rate and internet speed promote MOOC demand. In OECD countries, adult education level have positive and significant impact on MOOC demand. In China, we observe a positive and significant wage effect.

Our third essay relates to the development of human capital. We investigate the impact of family and teacher human capital/credentials on student ability development in middle schools in China. Our data are obtained from China Education Panel Survey (CEPS), which contain detailed information about student, homeroom teachers, major subject teachers, parents, and schools. Our ability measures include both cognitive and noncognitive ability. Cognitive ability is measured by a national test designed by CEPS, and our noncognitive ability measures include confidence, college intention, perseverance, and behavior problems. We find both family background and teacher human capital have significant impact on the development of cognitive ability. In addition, we observe teacher rank is the most predictive among teacher credentials. With respect to noncognitive ability, we family plays a more crucial role compared to teacher human capital.

CHAPTER 1

INTRODUCTION

Human capital may be defined as “the knowledge, skills, competence and other attributes embodied in individuals that facilitate the creation of personal, social, and economic well-being” (OECD, 2001). According to Coleman (1990), the formation of human capital theory is the most innovative and important development to education economics in the second half of 20^{th} century. According to the theory, in addition to natural resource and capital, human capital accounts for a significant amount of the wealth in the whole society. Specifically, human capital accounts for more than 50% of the national wealth except for Middle East counties that rely heavily on oil industry (World Bank 1997). In addition, people’s human capital such as skills, knowledge, and creative thinking have become essentially for themselves to make a living and for the development of the society.

Since the reform and open-up policy, China is experiencing a fast economic growth. Studies have shown that human capital plays an important role in such a fast growth, and it also narrows the economic gap between developed and less developed regions (Fleisher et al. 2009). Based on their calculation, for economic efficiency, it is important to put the investment in human capital as the first priority instead of the development of infrastructure.

Because of the importance of human capital, there are extensive literature about its development and its impact on individual academic and career outcomes (Herrnstein and Murray 1994, Heckman et al. 2006, Cunha and Heckman 2007). Traditionally, human capital is measured by education and on-the-job training (i.e., experience). Nowadays, as data become more available, more detailed measures of human capital such as cognitive and noncognitive abilities have drawn increasing attention. In literature, cognitive ability is commonly measured by IQ test, the Armed Forces Qualification Test (AFQT), and reading, writing, mathematics, and science test administered by educational institutions. Noncognitive ability is usually measured Rotter's measure of locus of control (Rotter 1966), the Rosenberg self-esteem scale (Rosenberg 1965), the Five-Factor Model of Personality (Muller and Plug 2006), and emotional intelligence (Goleman 2000).

Utilizing these new measures of human capital such as cognitive and noncognitive ability, Chapter 2 evaluates the relationship of human capital and leadership. Our data are obtained from the Programme for the International Assessment of Adult Competencies (PIAAC) survey. We define leaders as individuals who supervise more than five employees. To test for robustness, we also use other measures such as supervise more than ten employees or having a "manager" occupation. Our cognitive ability measures include numeracy, literacy, and problem solving, and our noncognitive ability measures include perseverance, openness to learning, and social trust. Our basic estimation is based on linear probability estimation, which is based on assumptions that our ability measures can adequately capture unobserved cognitive ability and both ability develop early and reach stability in late adult. Basic estimation results show that both cognitive and noncognitive are important predictors of leadership, and the most important ability measures are problem-solving and perseverance.

A further investigation of this question includes the possibility of imperfect ability measure and reverse causality, which will cause endogeneity issues. To deal with this issue, we apply three techniques. The first one is multiple indicator approach. We treat numeracy, literacy, and problem-solving as three indicators for ability and apply problem-solving as an indicator of cognitive ability and the other two measures as instruments. The second approach is IV estimation. We construct our first group of IVs based on language proficiency and cultural involvement, including whether parents were foreign born, whether the test language is the language mostly spoken at home, and the number of years spent in foreign countries. Our second group of IVs include whether participants were looking after children during survey or interrupted by other activities. The third approach is Control Function approach, which is more efficient when our endogenous variables include nonlinear terms. The results generated using these three techniques are consistent with our basic estimation in general. Thus, we conclude that cognitive and noncognitive abilities are important predictors of leadership.

Chapter 3 evaluates a new form of education, Massive Open Online Courses (MOOC), which has gained increasing attention since its appearance in 2012. Education is an important input of human capital. Many countries enforce a strict mandatory education system to make sure children in the country receive proper education. However, with the development of online education, more and more individuals have the option to take online courses, especially at the college level. To gain better understanding of online education, especially MOOC, we conduct a across study of OECD countries and a cross province study in China. We use search engine data to proxy MOOC demand. Specifically, MOOC demand in OECD is represented by search index provided by Google Trends, with the key words “MOOC+Coursera+Udacity+edX”. Similarly, MOOC demand in China is represented by search index provided by Baidu Index. We find that in both studies, higher unemployment rate and faster internet speed promote MOOC demand. In OECD countries, percent of college and high school graduates have significant impact on MOOC demand. In China, we observe a significant and positive wage effect. Notably, we do not observe any significant impact from tuition per capita, which indicates that there is no strong complement or substitute effects between MOOC and traditional education.

Chapter 4 evaluates the cognitive and noncognitive ability development among middle schools in China. Even though human capital has been proved to be an important reason for economic development in China, the research relates to ability development is limited. In this chapter, data are obtained from China Education Panel Survey (CEPS), a large-scale, nationally representative, longitudinal survey starting with the 7th and 9th graders in the 2013-2014 academic year. The survey conducts detailed surveys towards students, parents, teachers, and schools. By combing these information, we are able to identify the ability effect from family background and teacher credentials. In this chapter, cognitive ability is measured based on a national standard test, and noncognitive ability is constructed based on the information from surveys, including confidence, college intention, perseverance, and behavior problems. Family background variables include family environment such as parental relationship, study related materials/equipment at home, family size, parental education, and family income. Teacher credentials include rank, educational background, and years of experience. Our basic estimation result shows that both family background and teacher credentials are important predictors of cognitive ability. With respect to noncognitive ability, family plays a more critical role compared to teacher human capital.

As a future investigation to this issue, we also control for cross- and within-school sorting by incorporating school fixed effects and restricting our sample to classrooms with random assignments. In addition, we conduct the Hausman Taylor estimation to recover the coefficients of teacher human capital that does not change within class. The results are in general consistent with our previous basic estimation. Notable, we find teacher rank as the most effective predictor of teach credential.

This dissertation makes three contributions to the current literature. First, based on a new, nationally representative, and cross-country comparable sample collected by the OECD, Chapter 2 presents a thorough economic analysis of leadership as a labor market outcome from the perspective of human capital. Second, in Chapter 3, we fill the blank of MOOC-related research by conducting an economic analysis of MOOC demand among OECD countries and within China. The employment of search engine data and the proper choice of keywords can effectively proxy the actual MOOC demand in each geographic unit. Third, in Chapter 4, we add to the literature by evaluating both cognitive and noncognitive ability development among middle school students in China based on a new dataset from CEPS. By comparing the role of different teacher human capital measures, our study provides potential policy implications about the effectiveness of current teacher evaluation and promotion system in China.

The rest of the dissertation is organized as follows. Chapter 2 presents the relationship between human capital and leadership. Chapter 3 evaluates factors affect MOOC in OECD countries and in China. Chapter 4 studies the cognitive and noncognitive ability development among middle school students in China. Chapter 5 concludes.

CHAPTER 2

HUMAN CAPITAL AND LEADERSHIP: THE IMPACT OF COGNITIVE AND NONCOGNITIVE ABILITIES

2.1 Introduction

An integral characteristic of managers and supervisors is their leadership, which is essential for the development of businesses, governments, and other organizations. Despite the importance of leadership, very few economic studies have looked at the impact of human capital on leadership. In this study, we aim to fill the gap by conducting an economic analysis on leadership and use various human capital measures, including cognitive and noncognitive abilities, to investigate how they affect an individual's potential to become a leader.

Human capital has traditionally been measured through education and on-the-job learning (Mincer 1974). However, these indicators do not fully represent an individual's human capital. Other indicators of human capital, such as cognitive ability, have recently drawn more attention in studying individuals' development (Murnane et al. 1995, Cawley et al. 2001, Ree and Carretta 2002). Along with cognitive abilities, noncognitive abilities that represent personality, behaviors, and attitudes are also increasingly recognized as important factors in determining academic and career outcomes (Heckman et al. 2006, Lindqvist and Vestman 2011, Eren and Ozbeklik 2013). For instance, some individuals with a high level of intelligence do not succeed at work due to low levels of noncognitive abilities, such as a lack of persistence, reliability, or self-discipline (Heckman and Rubinstein 2001). Incorporating these ability measures into analysis can avoid the unobserved heterogeneity bias because abilities are likely to be correlated with other human capital measures such as education in the model.

In the current study, we utilize the newly available data from the Programme for the International Assessment of Adult Competencies (PIAAC) survey. We adopt three measures of cognitive ability (i.e., numeracy, literacy, and problem-solving abilities) and three measures of noncognitive ability (i.e., perseverance, openness to learning, and social trust). Our leadership analysis is based on a job-matching framework. That is, individuals with strong leadership are assigned to managerial level jobs. We adopt a number of alternative measures of leadership from various aspects in our empirical work, such as the complexity of jobs and occupation.

Our study has been inspired by the extensive research in organizational behavior and psychology. Dinh et al. (2014) reviewed existing leadership studies across 10 top-tier journals since 2000. Based on the review, we can see that studies on leadership vary dramatically. More specifically, some studies investigate how individual characteristics (i.e., traits) differentiate leaders from non-leaders (so-called “trait theory”), for example, Zaccaro (2007) and Judge and Bono (2000); and some studies focus on leaders’ behaviors (e.g., task-oriented or person-oriented) on the job and their relationships with managerial effectiveness (e.g., Carson et al. 2007); and some other studies investigate how a leader’s effectiveness is contingent upon various situational factors (e.g., Yukl 2008).

Our study is related to the literature about the effect of individual traits on leadership. We add to the short list of papers that examine an individual’s characteristics that foster him/her to emerge as a leader (Daly et al. 2015, Judge et al. 2010, Mumford et al. 2007, and Arvey et al. 2007). This study contributes to the literature in the following ways.

First, we conduct a thorough economic analysis of leadership as a labor market outcome from the perspective of human capital. We treat an individual's traits as a part of human capital and study how human capital affects one's chance to emerge as a leader. Personal traits are relatively stable and coherent integrations of personal characteristics. Certain traits are important for leadership, such as cognitive abilities, personality, motivation, social appraisal and interpersonal skills (Zaccaro et al. 2004, Mumford et al. 2007). We discuss in detail the relationship among those traits and traditional measures of human capital, such as education, and their impact on leadership.

Second, our study is based on a new, nationally representative and cross-nationally comparable sample collected in the PIAAC survey conducted by the OECD. PIAAC has been adopted by 33 countries in the latest survey in 2014. The data provide new measures of an individual's specific abilities. Therefore, we add to the literature of human capital by adopting a much more comprehensive measure of human capital, moving considerably beyond the traditional measures by incorporating measures of cognitive and noncognitive abilities.

Additionally, we specifically investigate a number of challenging issues, such as the potential endogeneity caused by the imperfect measurement of abilities and by potential reverse causality between measured abilities and leadership experience. We attempt to address these issues by exploring the data and using alternative estimation techniques under different assumptions about individuals' abilities.

In economic studies on leadership, although there are a few theoretical papers about the relationship between ability and job assignment (Rosen 1982, Gibbons and Waldman 2006), empirical studies on factors affecting the potential of being a leader from the human capital prospect are rare. Two economic studies are remotely related to our work here. Kuhn and Weinberger (2005) showed that people who held leadership positions in high school are more likely to earn higher wages and take managerial positions as adults. Yet, this study focused on leadership skills and wages. In another study, Borghans et al. (2008) studied the relationship between interpersonal styles (i.e. direct or caring) and job assignment. They showed that personality at age 16 can help predict job assignment in the future. For example, relatively caring (direct) people will end up working at jobs that require more caring (direct) personalities. Although it is not directly about leadership, it shows how personal traits affect one's future job.

This paper is organized as follows. In Section 2, we set up a simple theoretical framework on human capital and leadership. Section 3 introduces the data and ability measures. In Section 4, we discuss the relationships among different human capital measures and issues related to the nature of ability and its measurement. Section 5 presents basic results. In Section 6, we investigate various endogeneity issues, and Section 7 concludes.

2.2 A simple theoretical framework

Measuring leadership on a quantitative level is challenging because there are many definitions of leadership. In general, leadership contains a process of motivating and influencing the activities of other individuals or a group of individuals toward a common goal in a given situation (Hersey and Blanchard 1988). By developing missions, setting strategies, and motivating others, an effective leader disaggregates a complex project into relatively easy tasks and then assigns them to individual employees. In a firm, a straightforward demonstration of leadership is through job assignment. Thus, we apply a job-matching framework to model leadership. That is, in an efficient firm, individuals with good leadership are assigned to more complex jobs.

Our theoretical framework follows Gibbons and Waldman's (2006) conceptualization of job assignment and human capital. Assume that all firms have two kinds of jobs denoted by index j: managerial jobs ($\mathrm{j} = 1$), which are relatively more complex, and ordinary jobs ($\mathrm{j} = 0$), which are less complex. Since leadership can be revealed by job assignment, we define leaders as those individuals who conduct managerial jobs in a firm ($\mathrm{j} = 1$).

Job assignment is based on productivity. In a firm, if an individual is more productive in a managerial job than in an ordinary job (i.e., a demonstration of leadership), then the individual will be assigned to managerial level jobs. Specifically, we define the output of worker i assigned to job j as y_{ij},

$$y_{ij} = d_j + \mathrm{G}_j(S_i) + c_j h_i + u_{ij},$$

where d_j is the output of job j that is independent of the worker's human capital. S_i represents individual i's level of education. $\mathrm{G}_j(S_i)$ is a function of education, and the impact of education on the output depends on the individual's job assignment. $\mathrm{G'}_j(S_i) > 0$ for $\mathrm{j} = 0$ or 1, and h_i represents individual i's human capital in addition to education. We will refer to h_i as non-schooling human capital, including on-the-job learning ($Learn_i$), cognitive ability (Cog_i), and noncognitive ability ($NonCog_i$). In the above equation, c_j is the sensitivity of the output of job j to the individual's non-schooling human capital, and u_{ij} is the error term.

Job characteristics include c_j and d_j, and we assume $d_0 > d_1$ and $c_0 < c_1$. As can be seen, h^* is the level of non-schooling human capital at which a worker is equally productive at jobs 0 and 1 (i.e., $E(y_{i0}) = E(y_{i1})$). Given a fixed schooling level, Figure 1 displays the relationship between non-schooling human capital and output. We can show that

$$h_i{}^* = \frac{d_1 - d_0 + G_1(S) - G_0(S)}{c_0 - c_1}.$$

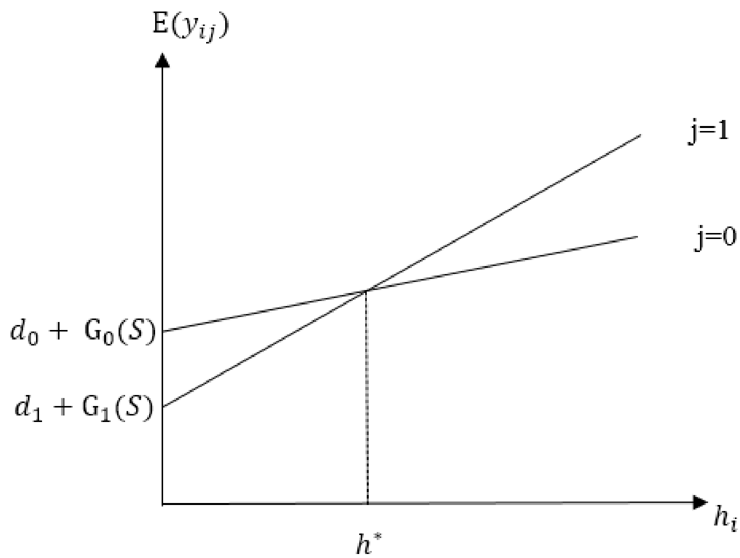

Figure 1. Non-schooling human capital and output, given a fixed education level

Given a fixed education level, when an individual has zero non-schooling human capital (i.e., $h_i = 0$), he/she will be relatively more productive at ordinary jobs that does not require intensive skills. That is, $d_0 + \mathrm{G}_0(S) > d_1 + \mathrm{G}_1(S)$. As h_i increases, the productivity growth in managerial jobs is relatively faster than in ordinary jobs ($c_1 > c_0$). To be a leader, in general, an individual must be more productive in a managerial job than in an ordinary job. As a result, an individual must have a non-schooling human capital higher than h^*. According to Figure 1, given $h_i > h^*$, at a particular school level, the higher the non-schooling human capital, the larger the difference between productivities between a managerial and an ordinary-level job, and thus the higher the probability of being a leader. If education level increases, the productivities are expected to increase (i.e., $\mathrm{E}(y_{ij})$ shifts upwards).

Based on the theoretical framework above, the probability of being a leader can be written as a function of human capital. That is,

$$Prob(leader)_i = f(S_i, h_i) + u_{ij} = f(S_i, Learn_i, Cog_i, Noncog_i\ ) + u_{ij}.$$

Because leadership can be reflected by the complexity of the job, in our estimation, we measure leadership in a number of alternative ways. More specifically, we use the number of employees an individual supervises to define leadership. Since a complex job requires more than one employees, the demonstration of leadership requires a group of individuals that are led by the leader. This measure can capture the core function of being a leader, and it can be made tighter or looser by specifying different number of supervisees. Additionally, we measure leadership based on occupation, i.e., manager position. One limitation for this measure is that in some industry, one can have a title as manager but in fact does not supervise others.

For human capital measures, traditionally, most studies have used education and on-the-job learning when evaluating the impact of human capital on individual career outcomes. As a result, an individual’s cognitive and noncognitive abilities, to the extent that they are not fully reflected by education and on-the-job learning, are often left in the error term. In this study, we include a number of cognitive and noncognitive ability measures, in addition to education and on-the-job learning.

Our model is estimated using both linear probability models (LPM) and the Probit model. The advantage of LPM is that it does not depend on the normality assumption of the underlying distribution of the error term to get a consistent estimation, and it is robust to heteroskedasticity.[1] We also apply the weighted least squares (WLS) estimation to obtain more efficient estimates of the LPM model.

One concern for estimating the model is about the ability measures. It is generally difficult to find a perfect measure of one's ability. If those measures cannot fully capture the true ability, then the ability measures themselves may be correlated with the "uncaptured" part of ability; and other regressors may also be correlated with the "uncaptured ability" left in the error terms. Both cases will result in omitted variable bias. Another concern is whether ability can be changed as time goes, for example, due to job experience and aging. If an individual's inner ability can be improved via on-the-job learning, then it is possible to have a reverse causality problem. For example, being in a leadership position may increase an individual's ability. We will investigate those potential issues in details below.

2.3 Data and ability measures

Our data originate from the Programme for the International Assessment of Adult Competencies (PIAAC) survey that was initiated by the OECD.[2] The PIAAC survey collects internationally comparable data about key cognitive and workplace skills of adults between the ages of 16 and 65 in OECD countries and other participating

[1] For the OLS type estimation, the normal distribution assumption for the error term is not needed for consistency and asymptotical normality.
[2] The link provides access to the PIAAC data: http://www.oecd.org/site/piaac/publicdataandanalysis.htm

countries.[3] The first round survey took place in 2012 and 24 countries participated. Nine additional countries participated in the second round in 2014. In each country, the PIAAC surveyed more than 5,000 individuals (the minimum response rate was 50%). Compared with other adult ability surveys, the PIAAC has several advantages: It provides recent and, thus, up-to-date data, covers a large number of countries, provides substantially larger sample sizes per country, and offers several in-depth measures of cognitive and noncognitive abilities.

In our study, we use United States survey data. We focus our study on the manufacturing, trade, and service industries.[4] Moreover, we include only full-time paid employees. We also limit the age range to 25 to 54 years in order to avoid including participants who are in the earliest stage of their career and those who are close to retirement. As a result, our final sample consists of 970 observations.

Our primary measure for leadership is defined based on number of individuals an individual supervises. This measure has the advantage of excluding individuals who have a manager title but do not actually supervise a team. Specifically, we define an individual as a leader if he/she supervises more than five employees and as a non-leader if the individual does not supervise anyone.[5] Based on this measure, we have 249 leaders and

[3] The PIAAC is different from the Programme for International Student Assessment (PISA) such that PISA measures the skills of 15-year-old students in mathematics, reading, and science, whereas the PIAAC measures the skills of adults in numeracy, literacy, and problem solving in technology-rich environments.
[4] Other industries that are included in the PIAAC survey are agriculture and the military. These industries involve different mechanisms of becoming a leader, and are not considered in the current study.
[5] Our leadership variable is based on two PIAAC questions: "Do you manage or supervise other employees?" and "How many employees do you supervise or manage directly or indirectly?" The response options for the first question include "Yes" and "No" and for the second question include "1 to 5 people," "6 to 10 people," "11 to 24 people," "25 to 99 people," and "100 or more people." If an individual responds "no" to the first question, then he/she will be treated as a non-leader. Note that we exclude individuals who supervise one to four employees in order to have a clear-cut difference between leaders and non-leaders.

721 non-leaders, with leaders accounting for 25.67% of the sample.[6] In addition, we construct two additional leadership measures: "Lead10" and "Manager." Lead10 defines leaders as those who lead more than ten employees. With this definition, we have 151 leaders and 721 non-leaders, with leaders accounting for 17.32% of the sample. In addition, we define a leader variable on the basis of the occupational classification of the respondent's job at the 2-digit level as defined by the International Standard Classification of Occupations (ISCO-08). If the occupation is "Manager," then the individual is defined as a leader.[7] With this definition based on "Manager," we have 159 leaders (12% of the sample) and 1,166 non-leaders. The correlation between Manager and Lead5 is 0.44, between Manager and Lead10 is 0.45, and between Lead5 and Lead10 is 1.00 (as Lead10 is a subset of Lead5).

On the ability measures, in the literature, commonly used measures of cognitive ability include, amongst others, IQ tests, the Armed Forces Qualification Test (AFQT), and reading, writing, mathematics, and science tests administered by educational institutions. The PIAAC survey collects information on each individual's cognitive abilities in three domains and provides three new measures on cognitive abilities: numeracy (*NUM*), literacy (*LIT*), and problem solving in technology-rich environments (*PS*).[8] Each of the three skill domains measures related and yet somewhat distinct

[6] In this study, we treat leader as a binary variable. It is possible to use the number of people that an individual supervises as a quantitative measure of leadership and then apply the Ordered Probit model. One concern with this approach is that the differences in the numbers of people supervises may not reflect the complexity of leadership.
[7] Individuals are defined as managers if they belong to one of these occupational groups: administrative and commercial managers (ISCO=12), production and specialized services managers (ISCO=13), hospitality, retail, and other services manager (ISCO=14).
[8] In the PIAAC, problem-solving ability is measured in technology-rich environments, representing a special type of problem solving. However, it still reflects the general problem-solving ability for leadership.

dimensions of an individual's skill set and is represented by a 500-point scale (ranging from 0 to 500) with higher points denoting a higher level of desirable skills.[9]

The definitions of each of the three domains provided by the PIAAC are as follows (OECD 2013):

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.[10]

Literacy: the ability to understand, evaluate, use, and engage with written texts to participate in society, to achieve one's goals, and to develop one's knowledge and potential.[11]

Problem solving (in technology-rich environments): the ability to use digital technology, communication tools, and networks to acquire and evaluate information, communicate with others, and perform practical tasks.[12]

The sample statistics for the three cognitive abilities are reported in Table 1. Leaders have higher average cognitive ability scores in all three domains than non-leaders. The average scores for leaders in numeracy, literacy, and problem solving are 2.78, 2.88, and 2.88, respectively, whereas the scores for non-leaders are 2.73, 2.86, and 2.81, respectively; and the difference in problem solving between leaders and non-leaders

---

[9] When presenting descriptive statistics and regression analyses, we divide all cognitive ability scores by 100. Thus, the numeracy, literacy, and problem solving scores range from 0 to 5 in the analyses that we apply.

[10] Numeracy tasks require, for instance, calculating the number of layers of tea candles packed in a box given other information or calculating the cost of a trip from a motor-vehicle logbook.

[11] The literacy test contains questions that require finding the right contact information in a simulated website, identifying the name of the author of a particular book in a simulated library website, and extracting certain information from given paragraphs or tables.

[12] Problem-solving questions include tasks such as reserving a meeting room on a particular date using a reservation system, organizing a family get together, and locating information on a spreadsheet and then e-mailing the requested information.

is statistically significant.[13] In addition to this, the standard deviations of the measured cognitive abilities of the leaders are smaller than those of the non-leaders, except for numeracy.

Similar to cognitive ability, there are many different ways to measure noncognitive ability. Some commonly used indices of noncognitive ability include Rotter's measure of locus of control (Rotter 1966), the Rosenberg self-esteem scale (Rosenberg 1965), the Five-Factor Model of Personality (Muller and Plug 2006), and emotional intelligence (Goleman 2000). For instance, Heckman et al. (2006) use Rotter Locus of Control Scale and Rosenberg Self-Esteem as noncognitive ability measures. The Rotter scale measures the degree to which individuals feel they control their life and Rosenberg scale measures perceptions of self-value.

[13] The average scores for the three measures in the whole sample (i.e., 970 observations) are 2.74, 2.86, and 2.83, with standard deviations of 0.47, 0.41, and 0.41, respectively.

Table 1. Descriptive statistics

	Lead5=1 (Obs=249)			Lead5=0 (Obs=721)			Difference
Variable	Obs.	Mean	SD	Obs.	Mean	SD	
Cognitive skills							
Numeracy	249	2.778	0.468	721	2.729	0.467	0.049
Literacy	249	2.875	0.397	721	2.861	0.412	0.014
Problem solving	249	2.882	0.399	721	2.811	0.413	0.071*
Noncognitive skills							
Perseverance (yes=1)	249	0.321	0.468	721	0.239	0.426	0.082*
Openness to learning (yes=1)	249	0.470	0.50	721	0.452	0.498	0.018
Social trust (yes=1)	249	0.309	0.463	721	0.264	0.441	0.045
Education degree							
Below high school	249	0.056	0.231	721	0.031	0.172	0.025
High school	249	0.422	0.495	721	0.490	0.50	0.068*
Bachelors	249	0.313	0.465	721	0.295	0.457	0.018
Masters or Ph.D.	249	0.209	0.407	721	0.184	0.388	0.025
Others							
Male (yes=1)	249	0.570	0.496	721	0.449	0.498	0.121*
Experience (years)	249	20.133	8.450	721	18.031	8.824	2.102*
Married (yes=1)	213	0.817	0.388	577	0.730	0.445	0.087*
Either parent had a college degree (yes=1)	243	0.407	0.492	694	0.451	0.498	0.044
Number of children	249	1.731	1.213	721	1.422	1.298	0.309*
Industry Dummies							
Manufacturing (yes=1)	248	0.165	0.372	721	0.215	0.411	0.050*
Trade (yes=1)	248	0.278	0.449	721	0.151	0.358	0.127*
Service (yes=1)	248	0.556	0.498	721	0.634	0.482	0.078*

Note: 1. Our final sample of 970 included only individuals with all three cognitive ability measures.
2. When presenting descriptive statistics and regression analyses, we divide all cognitive ability scores by 100. Thus, the numeracy, literacy, and problem solving scores range from 0 to 5 in the analyses that we apply.
3. * indicates that the difference between leader and non-leader was significant at the 10% significance level.

In our study, based on the information available in the PIAAC data, we construct three noncognitive ability measures including perseverance, openness to learning, and social trust. They are reflected in the five-factor model of personality, which includes extraversion, agreeableness, conscientiousness, and openness to experience, and are likely to be important for leadership. More specifically, perseverance (i.e., maintaining high energy levels even in difficult circumstances) is an essential trait for leaders (Dries and Pepermans 2012). In addition, to be better prepared for unexpected new situations, leaders should keep learning new things to expand and deepen their knowledge base. Social trust helps build a comfortable relationship between a leader and team members. Moreover, a leader who trusts team members can also gain their trust relatively easier, and a trust relationship is important in leadership effectiveness (Dirks and Ferrin 2002).

In our data, perseverance is measured with the question "I like to get to the bottom of difficult things." The respondents select their response from a 5-point scale with the anchor points "not at all," "very little," "to some extent," "to a high extent," and "to a very high extent." In order to be categorized in the high-perseverance group, the response has to be "to a very high extent."[14] Thus, if a person responds to the statement "to a very high extent," and none of the situations listed in Footnote 14 occur, we set *Perse=1.* Otherwise, *Perse=0.* We place a strict rule on defining a high level of perseverance because the responses per se are subjective, and a person may be more likely to be overly positive in a self-evaluation. Similarly, openness to learning is measured by the response to the statement "I like learning new things." We set *Learn=1* if an individual responds "to a very high extent." Otherwise, *Learn=0.* Social trust is measured with two statements: "There are only a few people you can trust completely"

[14] Due to the highly subjective nature of responses to this question, we double check the individual's choice with additional information on the actions and activities occurred during the interview: (1) the respondent held a conversation with someone else in the house besides the interviewer; (2) the respondent engaged in domestic tasks such as washing or cooking; (3) a television set, radio, game console, or music player was in use in the immediate vicinity of the respondent. We argue that people with strong perseverance will not conduct the above activities during the interview.

and “If you are not careful, other people will take advantage of you.” Responses to the statements are again measured on a 5-point scale with the anchor points “strongly agree,” “agree,” “neither agree nor disagree,” “disagree,” and “strongly disagree.” We define that individuals have good social trust (i.e., *Strust=1*) if they answer “disagree” or “strongly disagree” to at least one of the statements; otherwise, *Strust=0*.

Table 1 presents the descriptive statistics for the noncognitive abilities of leaders and non-leaders. Among leaders, individuals with strong perseverance, openness to learning, and good social trust account for 32%, 47%, and 31%, respectively. The percentages drop to 24%, 45%, and 26% for non-leaders. The largest difference in percentage points is observed in perseverance. It is likely that leaders generally need a high level of perseverance to solve complex problems when completing their work duties.

Education is measured by individuals’ highest academic degrees such as high school diplomas, bachelor’s, master’s, and Ph.D. degrees. Similar to the patterns of cognitive and noncognitive abilities, leaders have higher levels of education than non-leaders. Among leaders, 52% have bachelor’s degrees or higher, whereas, among non-leaders, only 48% have a similar level of education.

The survey data do not provide the exact age of the respondents but report in 5-year age categories. We estimate the age of the individuals by taking the median of these categories. Further, we define experience as age minus years of schooling minus six.[15] Table 1 shows that leaders generally have more years of work experience, and men are more likely to be a leader (57% of the leaders). The family background variables include

[15] This calculation provides an upper bound for experience because we do not have information on unemployment, and will not cause endogeneity for leaders and non-leaders.

marital status, number of children, and parental education. If one of the parents receives tertiary education, then parental education equals one; otherwise, it equals zero.

2.4 Further discussions on human capital measures

Relationship among human capital measures

We expect that the sets of human capital measures would be related to each other. A comprehensive analysis of such relations is integral to the understanding of their impact on leadership. Table 2 shows that the cognitive ability measures are highly correlated with each other (all correlations > .82). The correlation between numeracy and literacy is the highest. It is possible that comprehending a numeracy task requires good literacy. Table 2 further indicates that cognitive and noncognitive abilities are also positively correlated, but the correlations are much smaller (all correlations < .20) than those within the cognitive abilities.

Table 2. Correlations between abilities

	Numeracy	Literacy	Problem solving	Perseverance	Openness to learning	Social trust
Numeracy						
Literacy	0.886*					
Problem solving	0.824*	0.858*				
Perseverance	0.097*	0.061*	0.065*			
Openness to learning	0.043	0.057*	0.066*	0.354*		
Social trust	0.153*	0.154*	0.187*	-0.011	0.141*	

Note: The numbers are correlation coefficients between different ability measures. * indicates the correlation was significant at the 10% significance level.

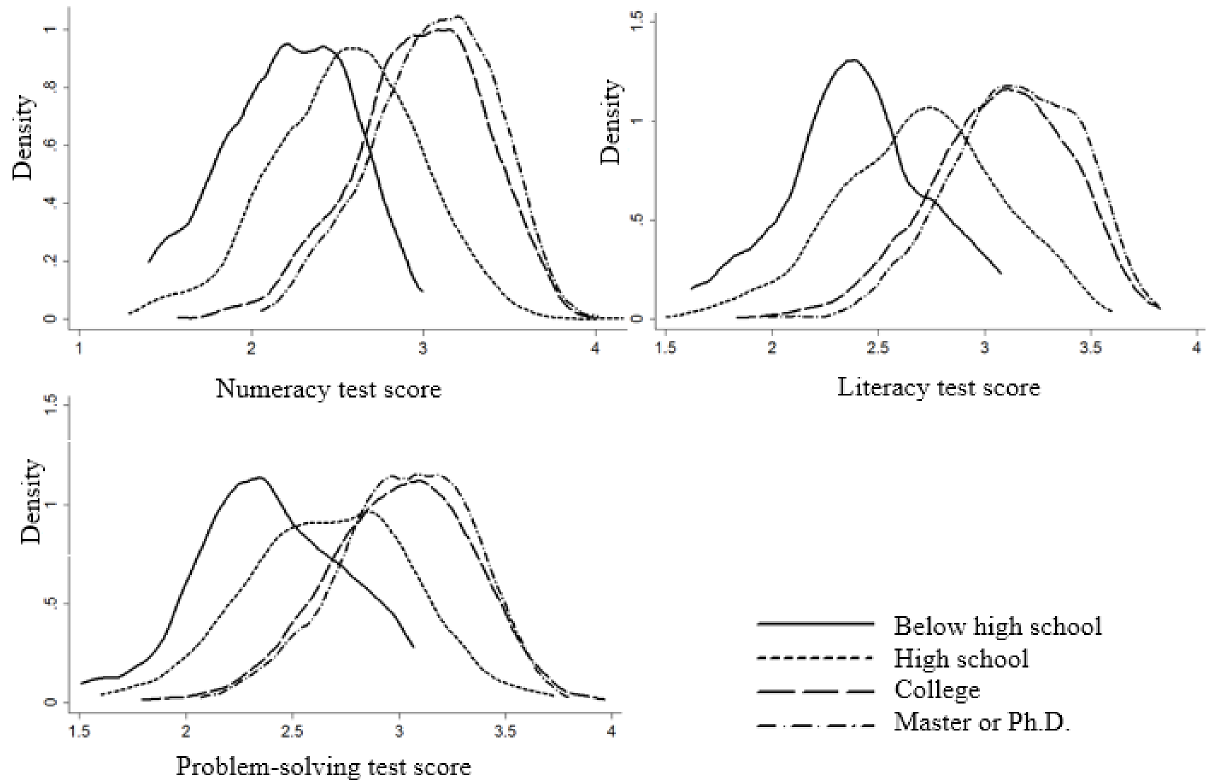

Figure 2. Kernel distribution of the three cognitive abilities by education levels

Figure 2 depicts the distributions of the three cognitive ability measures separated by education level. Clearly, as education level increases, the mean ability levels increase for all cognitive ability measures as well. On average, individuals with master's or Ph.D. degrees have the highest abilities, and those with no high school diploma have the lowest. The largest increase is observed for numeracy. Individuals with master's or Ph.D. degrees have a 38.7% higher average score in numeracy than those with less than a high school education. It is interesting that problem-solving scores show the smallest gap between different educational levels.

For noncognitive abilities, as shown in Table 3, the percentage of individuals with good noncognitive abilities is higher for individuals with college or graduate degrees compared with those with a high school education or below. For instance, for those with graduate degrees, the proportion with high perseverance is 8.1 percentage points higher than for those with less than a high school education. A similar pattern is found for openness to learning, where the difference between graduate education and below a high school education is 10 percentage points. Social trust also shows a positive relation to education, and the gap between the higher and lower educational levels is even larger. According to Hooghe et al. (2012), individuals with an attitude of trust are likely to develop good social relationships and a clear academic orientation, which in turn leads to higher education.

Table 3. Noncognitive ability and education level

Education	Perseverance	Openness to learning	Social trust
Below High School	23.5%	45.1%	17.6%
High School	21.5%	41.7%	20.2%
College	32.1%	52.0%	37.0%
Masters or Ph.D.	31.6%	55.1%	39.9%

Note: The percentages represent proportions of individuals with high noncognitive abilities (i.e., perseverance=1 based on the definition explained in the text) at each education level.

Imperfect measurement of abilities

In general, the three new measures of cognitive abilities adopted by the OECD are comprehensive. They are considered to be adequate measures of one’s cognitive ability. In this case, there will be no complications in estimating the leadership model. The corresponding results will be discussed in Section V.

However, due to the highly complex nature of human intellect, it is possible that the cognitive ability measures from the PIAAC still do not capture the entire range of human ability. For instance, cognitive ability is assumed to be composed of a number of hierarchically ordered abilities with a very general cognitive ability factor at the top of the hierarchy (McGrew, 2009). Numeracy, literacy, and problem solving are probably strongly related to this general factor, but they do not capture all aspects of human intellect. If so, there will still be uncaptured cognitive ability in the error term. If the three OECD measures are not correlated with the “remaining portion” in the error term, there are no additional concerns. Otherwise, those measures might still be correlated with “uncaptured” cognitive ability and will result in an endogeneity problem. We apply the Multiple Indicator approach to solve this problem in section VI.

On the other hand, noncognitive ability measures one’s personality. Based on the five-factor model of personality, an individual’s personality can be classified into five categories including extraversion, agreeableness, conscientiousness, openness to experience and neuroticism. Our three measures of noncognitive ability (i.e., perseverance, openness to learning and social trust) can fit into the first four factors as a representative of each category. They are unlikely to be correlated with the remaining noncognitive ability, neuroticism.

In addition, according to our data (Table 2), the correlation among our noncognitive ability measures are low, from -0.011 to 0.354. Heckman et al. (2006) also showed a smaller correlation among noncognitive abilities measures. Therefore, even if our noncognitive measures do not capture all noncognitive abilities, it is less of a concern that the measures included in the model are correlated with unmeasured noncognitive abilities in the error term.

Stability of ability and reverse causality

For the reverse causality issue, it critically depends on whether an individual's ability is stable during their lifetime. In general, both cognitive and noncognitive abilities develop early in life and reach stability in adolescence or early adulthood at latest. Thus, those abilities can be considered as antecedents of later outcomes including occupational status and leadership position. There might be an effect of the opposite direction (i.e., the level of cognitive and noncognitive ability is slightly influenced by the unique experiences leaders have), but this reversal effect should be small.

For cognitive abilities, a large number of studies in physiology have shown that intelligence exhibits an impressive level of stability from childhood throughout adulthood all the way to old age (e.g., Deary et al. 2000, Schalke et al. 2013). Based on Cuhan and Heckman (2007), individuals' cognitive ability becomes stable after age 10 or so, and they showed that early interventions were much more effective than later interventions. Regarding the noncognitive abilities, Cobb-Clark and Schurer (2012) showed the stability of the Big Five traits of personality. Specht et al. (2011) reported rather high levels of stability for several aspects of personality, in particular between the fourth and the sixth decade of life. Additionally, the empirical findings in Kuhn and Weinberger (2005) and Borghans et al. (2008) about the impact of noncognitive ability at a relatively young age (i.e., high school or age 16) on later career outcomes in adulthood also indicated a substantial level of stability of noncognitive abilities.

Moreover, in order to see how stable our ability measures are, we check how these measures change with age and job experience in our data. Table 4 presents the OLS estimation results with ability measures as dependent variables and age and experience group dummies as independent variables, separately. We divide our sample into six age groups and seven experience groups with five-year intervals. For age groups, the reference age group is 25-29 years old, and for experience groups, the reference group is 0-9 years of experience. As can be seen from Table 4, in general, both cognitive and noncognitive abilities exhibit a clear decreasing trend as age and experience increase.

Therefore, our data do not support that those measured ability scores increase with job experience.[16] Moreover, in our sample, the average age is 39, which is well beyond the early stage of development. Thus, it is reasonable to assume that an individual's inner ability remains rather stable throughout the age range (25-54 years old) we investigate and the potential reverse causality should not be a big concern in this study.[17] However, to further test this issue in our estimation, we explore the available information from the data and employ Instrument Variable approach to estimate our models. The results and discussions are presented in Section VI.[18]

---

[16] It is possible, though, the aging or cohort effect offsets the on-the-job learning effect, and thus we see a decreasing trend.

[17] The impact of on-job learning on one's productivity found in the literature may just reflect the increasing efficiency in job-specific tasks, instead of inner ability.

[18] Because of data limit, we test the reverse causality for cognitive abilities but not for noncognitive abilities.

Table 4. Change of ability over age and experience

	(1) Numeracy	(2) Literacy	(3) Problem solving	(4) Perseverance	(5) Openness to learning	(6) Social trust
Age						
D30-34	-0.146***	-0.140***	-0.143***	-0.027	0.013	-0.069
D35-39	-0.065	-0.088**	-0.077**	-0.019	-0.103**	-0.056
D40-44	-0.097**	-0.083**	-0.142***	0.005	-0.053	-0.006
D45-49	-0.180***	-0.154***	-0.207***	-0.020	-0.114**	-0.048
D50-54	-0.143***	-0.171***	-0.271***	-0.015	-0.156***	-0.025
	Numeracy	Literacy	Problem solving	Perseverance	Openness to learning	Social trust
Experience						
D10-14	-0.145***	-0.147***	-0.133***	-0.046	-0.035	-0.091**
D15-19	-0.164***	-0.167***	-0.139***	-0.044	-0.121***	-0.050
D20-24	-0.167***	-0.140***	-0.188***	-0.025	-0.126***	-0.033
D25-29	-0.218***	-0.190***	-0.243***	-0.005	-0.117***	-0.045
D30-34	-0.295***	-0.306***	-0.353***	-0.072*	-0.232***	-0.077*
D35-39	-0.600***	-0.652***	-0.803***	0.132	0.006	-0.332***

Note: 1. This table presents OLS estimation coefficients with ability variables as dependent variables and age and experience group dummies as independent variables.
2. For age groups, the reference age group is 25-29 years old. For experience groups, the reference group is 0-9 years of experience.
3. *** denotes significance at 1%, ** denotes significance at 5%, and * denotes significance at 10%.

2.5 Basic Results

In our basic estimation, the ability measures adopted by the OECD are assumed to be adequate for capturing one's unobserved ability, or they are not correlated with the "uncaptured" ability left in the error term. The basic estimation results are reported in Table 5.[19] As cognitive ability is one of the most frequently studied leader attributes, we first include the three cognitive ability measures in model 1. The literacy component of

[19] For models based on LPM estimation in Table 5, the majority of the predicted probabilities fall between zero and one. For instance, in our baseline model 2, there are eight predicted values smaller than zero (i.e. 1% of the sample) and no predicted values larger than one. This is an indication that the LPM model works well.

cognitive ability is specified as that its potential peaking effect on leadership is in its midrange. We conjecture that individuals who have very high literacy scores may also have personality traits associated with exceptional erudition that mitigate against their being chosen as leaders.[20]

Model 1 shows that out of the three cognitive ability measures, only problem-solving ability is positively and significantly related to leadership.[21] The coefficient of problem solving is 0.306. This indicates that when problem-solving score increases by 1 unit, the probability of being a leader increases by 0.306 (i.e., 30.6 percentage points). Because a 1-unit increase represents a 100-point increase in the problem-solving test score, it then means that a 10-point increase in the problem-solving score increases the probability of being a leader by 3.06 percentage points.[22]

[20] For numeracy and problem solving, we believe this is less likely to be an issue. We try to add the quadratic terms of numeracy and problem solving in regression. Both linear and quadratic terms become insignificant and other results remain robust.

[21] We also run this model with cognitive ability measures added to the model separately. We find that education is always a significant predictor of being a leader. In addition, problem-solving skills remain positive and significant, whereas numeracy and literacy remain insignificant in these models.

[22] We also include interaction terms between problem-solving ability and industry dummies in the analyses. None of these interaction terms is significant.

Table 5. Ability and leader

	Lead5				Lead10	Manager
	(1) LPM	(2) LPM	(3) PROBIT	(4) WLS	(5) LPM	(6) LPM
Numeracy	-0.031	-0.040	-0.046	-0.033	0.013	0.108**
Literacy	0.153	0.166	0.177	0.145	-0.012	-0.333
Literacy^2	-0.068	-0.068	-0.068	-0.053	-0.067	0.032
Problem solving	0.306***	0.295***	0.300***	0.273***	0.438***	0.122**
Education	0.093**	0.085**	0.083**	0.072**	0.113***	0.105***
Experience	0.006***	0.006***	0.006***	0.005**	0.006***	0.004***
Perseverance		0.090**	0.090**	0.084**	0.068*	0.023
Openness to learning		-0.027	-0.028	-0.032	-0.011	-0.001
Social trust		0.034	0.035	0.038	0.011	0.005
Male	0.097***	0.093**	0.098***	0.090***	0.028	0.009
Married	0.057	0.057	0.067	0.048	0.011	0.025
Either parent had a college degree	0.002	0.001	-0.004	-0.012	0.013	-0.010
Number of children	0.021	0.022	0.020	0.022*	0.021*	0.006
Constants	-0.667	-0.658	-3.644*	-0.639	-0.779	-0.028
Industry dummies	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	758	758	758	758	676	1040
R^2	0.075	0.083		0.106	0.126	0.066
Adjusted R^2	0.06	0.064		0.088	0.106	0.052
F-statistics	5.248	4.752		5.862	5.651	4.291
Pseudo R^2			0.072			

Note: 1. The Probit model presents the marginal effects evaluated at the means of the covariates.
2. *** denotes significance at 1%, ** denotes significance at 5%, and * denotes significance at 10%.

Problem-solving ability has been considered an important precursor of lifelong learning and later success in life.[23] Leadership comes along with the need for problem solving on a regular basis at different occasions and in a variety of circumstances. The

[23] For this reason, problem solving was included as a transversal skill in the arguably most important international large-scale educational assessment ever conducted, the Programme for International Student Assessment (PISA). In its 2012 cycle, it involved 15-year-old students in over 40 countries (OECD 2014).

problems a person faces at work are often complex or non-transparent, involve numerous constraints, and consist of large sets of variables. A high level of problem-solving ability can help leaders define exactly what the problem is and help them generate appropriate solutions to the specific problem at hand (Mumford et al. 2000). Connelly et al. (2000) found that complex problem-solving skills, together with social judgment skills and leader knowledge, accounted for significant variance in leadership even after controlling for general intelligence, motivation, and personality.

In addition, model 1 shows that if an individual has a college degree, his or her probability of being a leader increases by 9.3 percentage points. Experience also has a significant positive impact on leadership. This finding is in line with previous findings on the relation between experience and leadership. Using a sample of about 5,000 graduates of the Stanford MBA program in the late 1990s, Lazear (2010) showed that the number of previous positions held by individuals significantly increased their probability of being a manager.

Model 1 also shows that, after controlling for the human capital measures, males have a nearly 10 percentage point higher chance of being a leader than females. Parental education does not seem to influence a person's chance of being a leader, and marital status and number of children do not show any significant effects either.

After evaluating the impact of the cognitive ability measures, we then add noncognitive ability measures in model 2. The inclusion of noncognitive ability solves the omitted ability problem to a large extent as some human capital variables that are included in model 1 might have been correlated with noncognitive ability measures (e.g., perseverance). Out of the noncognitive ability measures, only perseverance shows a positive and significant effect.[24] More specifically, a high degree of perseverance increases the probability of being a leader by 9.0 percentage points. Perseverance (i.e., maintaining high energy levels even in difficult circumstances) is an essential trait for leaders. Leaders, who oftentimes face unexpected difficulties, obstacles, and discouragement, need strong perseverance to lead the team and work toward the goal that is to be achieved.

The results for education and the other human capital measures are similar to those for Model 1. However, due to the inclusion of additional human capital variables, the magnitudes of the coefficients are somewhat smaller. This is consistent with our hypothesis that the noncognitive abilities might have been positively correlated with education and problem-solving skills and thus may have resulted in a positive omitted variable bias when omitted. Other studies (e.g., Segal 2013) reported that a significant portion of the impact of cognitive ability on earnings could be attributed to noncognitive ability.

Models 3 to 4 present the results of the alternative estimation methods. The results of the Probit estimation in Model 3 are consistent with the OLS estimation of the LPM model (Model 2). In fact, the magnitudes of the marginal effects for the significant variables are almost identical to those from the LPM estimation. The human capital measures, including problem solving, perseverance, education, and experience, remain significant. The WLS estimation (Model 4), which is presumed to yield more efficient estimates than OLS estimation, exhibit results that are similar to the OLS estimates as

[24] If we include the noncognitive measures in the model separately, the results are similar. That is, perseverance is the only significant predictor out of the noncognitive abilities.

well, although the marginal effects for almost all variables are somewhat smaller on a descriptive level.[25]

In order to provide an additional check of the robustness of our results, we run models using the different definitions of being a leader: supervising at least ten people (Lead10 in Model 5) or being a manager (Manager in Model 6). With the somewhat stricter definition of leadership (i.e., leading more than ten employees), the results of the LPM remain generally consistent. However, in this alternative model, the effects of most human capital measures become larger and more pronounced, which is to be expected given that we are more selective in defining what constitutes a leader and what does not. In particular, the coefficient of problem solving increases to 0.438, a marked increase of 48% compared with the model with a more lenient definition of leadership (i.e., supervising more than five employees). Perseverance is still the only significant noncognitive ability measure, but its absolute impact is somewhat smaller.

When a leader is defined by the occupation “Manager,” the results are again generally consistent with those that are obtained with the two other definitions of leadership, Lead5 and Lead10, and, generally speaking, exhibit the same pattern of results even though some coefficients change in magnitude. These changes in the sizes of the coefficients are mixed; for example, the effect of education becomes larger, whereas the effects of problem solving and experience become smaller. A notable change is observed for numeracy, which is positive and significant only when “Manager” is used as the dependent variable. More specifically, if the numeracy scores increase by 10 points, the probability of being a manager increases by 1.08 percentage points, so even in this

[25] For WLS, in order to obtain more efficient standard errors, $1/(\hat{y}*(1-\hat{y}))$ is used as the weight, where $\hat{y}$ is the predicted value of the corresponding OLS estimation.

model, the effect remains rather small. In addition, perseverance becomes insignificant. The explanation for these changes is that the position of a manager can vary a great deal in terms of the number of people the manager supervises. Thus, the "Manager" definition is relatively loose in defining a leader, leading to a less distinct pattern of results than found in the previous models.

2.6 Imperfect measurement of ability and reverse causality

As discussed above, if the ability measures do not capture the entire range of one's inner ability and are correlated with the "uncaptured portion" of the ability, we will have an endogeneity problem. In order to investigate this potential problem, we treat all three cognitive measures as indicators of inner ability, and then apply the Multiple Indicator (MI) estimation method. In this case, those measures do not have to perfectly capture the true ability but only need to be indicators of it.[26]

Because we have multiple indicators of one's inner ability, numeracy, literacy, and problem solving, we can treat one as the indicator and others as instruments in the estimation. In theory, the roles of all three indicators, numeracy, literacy, and problem solving are interchangeable in the MI estimation. Based on estimation results in the above section, we can see that problem solving provides a relatively more comprehensive measure of the underlying cognitive ability. Therefore, we use problem-solving ability as an indicator of cognitive ability and the other two measures as instruments. It is clear

---

[26] In the indicator approach, suppose the original equation is $\mathrm{y} = \alpha_0 + \mathbf{x\beta} + \gamma \mathrm{q} + \mathrm{v}$, where q represents the omitted ability variable. Suppose we have multiple indicators of q from q_1 to q_n, indicator q_1 can be written as $q_1 = \delta_0 + \delta_1 q + a_1$, where $\mathrm{cov}(\mathrm{q}, a_1) = 0$ and $\mathrm{cov}(\mathbf{x}, a_1) = \mathbf{0}$. If we rewrite q as a function of q_1 and then substitute it back into the original equation, we then have $\mathrm{y} = \alpha_1 + \mathbf{x\beta} + \gamma_1 \mathrm{q}_1 + \mathrm{v} - \gamma_1 a_1$, where $\gamma_1 = \frac{\gamma}{\delta_1}$. If we express each of the rest of the indicators (i.e., q_2 to q_n) as a function of q and also assume the error terms are not correlated with a_1, then q_2 to q_n become valid instruments for q_1.

that numeracy and literacy are correlated with the indicator, problem-solving ability score, and thus can serve as instruments.

The results of the MI estimation are reported for Model 1 in Table 6. MI estimation solves the omitted ability bias caused by imperfect ability measure. The results are in line with the LPM estimation for Model 2 in Table 5. The impacts of education and experience become smaller, which, as expected, indicate that they have positive relationships with the omitted ability. Specifically, the coefficient of education drops from 0.085 to 0.075, and the coefficient of experience drops from 0.006 to 0.004. The impact from perseverance becomes larger, with an increase from 0.090 to 0.095. Problem solving becomes insignificant. A possible reason for the insignificance is the relative inefficiency of the IV estimation compared to the OLS estimation. Overall, the MI estimates do not contradict with the LPM estimation above.

Table 6. Multiple indicator, instrumental variable, and control function estimation

	(1) MI	(2) IV	(3) CF
Numeracy		-0.745	-0.727
Literacy		2.305	1.422*
Literacy^2		-0.234	-0.069
Problem solving	0.006	-0.459	-0.454
Education	0.075*	0.503*	0.482*
Experience	0.004**	0.005	0.005
Perseverance	0.095**	0.094*	0.094*
Openness to learning	-0.027	-0.052	-0.052
Social trust	0.045	0.034	0.022
Constant	-0.014	-1.582	-0.478
Number of observations	758	751	751
R^2	0.066	-	0.089
F-statistics	-	-	3.955
Endogeneity test (F-test)		1.202 (p=0.307)	1.44 (p=0.219)
Over-identification test (Chi^2)		0.022 (p=0.989)	

Note: *** denotes significance at 1%, ** denotes significance at 5%, and * denotes significance at 10%.

Additionally, we further investigate the potential reverse causality issue discussed above and apply the Instrument Variable estimation. We explored information from the PIAAC background questionnaire and constructed instruments for the endogenous variables. More specifically, we treat education and all cognitive ability measures as endogenous, caused either by imperfect measurement, i.e., they may be correlated with the remaining portion ability, or reverse causality, i.e., they may be changed as a result of leadership experience. The advantage of IV estimation is that it can solve both problems above, but (unlike the MI approach) it requires instruments outside those measurements. The instruments are required to be correlated with education and measured cognitive abilities in the model but not the model error term.

The first group of IVs are related to language proficiency and cultural involvement. Specifically, the IVs are: (1) whether both parents were foreign born, (2) whether the test language is the same as the language usually spoken at home, and (3) the number of years spent in foreign countries before immigrating to the United States.[27] These IVs might have affected the test scores and the level of education. However, they should be exogenous to the individual's inner ability.[28] The IVs in the first group are correlated, but the correlations are not high. For example, the highest correlation occurs

[27] Literacy enters into the model along with its squared term, which is also endogenous. In theory, we could include all quadratic terms and interaction terms of all instruments as additional instruments for this squared term. However, because most of the instruments are dummy variables, we include only one interaction term as an additional instrument, which is the interaction between "years spent in other countries" and "whether both parents were foreign born." Such an interaction can capture the difference under the situation that both parents were born outside the US but the years the individual spent in other countries are different. Such a difference is likely to affect people's language proficiency and cultural engagement and subsequently their test scores and education.

[28] It is possible that some of these instruments will be correlated with other leadership related factors in the error term, such as discrimination. We conjecture that such an effect should be small in the US, if any, especially when literary, problem-solving ability, and social trust have been controlled for in the model.

between whether both parents were foreign born and years spent in a foreign country (r = 0.730).

The second group of IVs are constructed on the basis of whether the participants were interrupted by any other activities during the survey, including (1) whether an individual was looking after children and (2) whether an individual was interrupted by some other activity, task, or event. These activities might have distracted individuals from concentrating on the tests, and, thus, might have negatively affected their test scores. Because these interruptions are likely to be beyond the individual's control, they are exogenous to the person's real ability and leadership related factors in the error term.

The descriptive analyses of the IVs indicate that individuals whose parents were born in foreign countries account for 15% of the sample, and those who usually speak the test language at home account for 93%. The average number of years spent in foreign countries before immigrating to the United States is 2.24. In addition, 10% of the participants were looking after children during the interview, and 11% were interrupted by some other activity, task, or event.

Appendix A lists the first stage estimation results, and the results indicate that the IVs are correlated with the endogenous variables. In particular, if English (the test language) is the language usually spoken at home, the scores for all three cognitive abilities are generally higher. This holds also for the level of education (i.e., attending college). The interaction between foreign-born parents and years spent outside the US before immigration is negative and significant for literacy, problem solving, and education.

The results of the IV estimation are reported for Model 1 in Table 6. The over-identification test could not be rejected, indicating that there is no evidence against the validity of our instruments. Consistent with the previous estimations, the IV approach shows that education and perseverance are significant. In particular, the probability of being a leader increases by 0.503 if an individual has a college degree (or a higher degree) relative to those without a college degree. This coefficient is much higher than the coefficient in our previous estimation of 0.085 (Table 5, Model 2).[29] In addition, the results of the IV analyses indicate that individuals with strong perseverance have a higher probability of being a leader of 0.094, which is similar to the previous LPM estimation of 0.090 (Table 5, Model 2). In the IV estimation though, problem-solving ability does not show a significant impact on leadership.

Considering that the endogenous variables include one nonlinear term (i.e., the squared literacy term), we also apply the control function (CF) estimation to further improve efficiency. When there are nonlinear endogenous regressors, CF is generally more efficient than IV estimation.[30] The CF estimation results are reported for Model 2 in Table 6. In general, they are similar to the pattern of results for the IV estimation with regard to education and perseverance, both in sign and magnitude. The CF further shows that literacy has positive and significant effects on being a leader.

---

[29] Many studies dealing with omitted ability bias, especially those based on the Mincer equation, have found that IV estimates are much higher than OLS estimates, although they are supposed to be smaller due to the removal of the positive ability bias. A common explanation is that the attenuation bias caused by measurement error dominated the omitted ability bias (Card 1999). Our results here seem to be in line with this finding.

[30] In the CF approach, we employ the estimated equation $y_1 = \boldsymbol{z_1\delta_1} + \alpha_1 G(y_2) + u_1$, where y_1 is the dependent variable, y_2 is the endogenous variable, $G(y_2)$ is a function of the endogenous variable, $\boldsymbol{z_1}$ is a $1 \times L_1$ sub-vector of $\mathbf{z}$, and $\mathbf{z}$ is a $1 \times$ L vector of exogenous variables. The reduced form of y_2 is $y_2 = \boldsymbol{z\pi_2} + v_2$. Assume $\mathbf{z}$ is independent of u_1 and v_2, $\mathrm{E}(y_1|\boldsymbol{z}, y_2) = \boldsymbol{z_1\delta_1} + \alpha_1 G(y_2) + E(u_1|\boldsymbol{z}, y_2) = \boldsymbol{z_1\delta_1} + \alpha_1 G(y_2) + E(u_1|\boldsymbol{z}, v_2) = \boldsymbol{z_1\delta_1} + \alpha_1 G(y_2) + \rho_1 v_2$. Then, the model could be estimated by OLS with v_2 added, whereas v_2 could be estimated with the residual from the reduced form of y_2 (i.e., $\widehat{v_2}$).

Overall, the IV estimation results do not contradict with basic results from OLS and Probit estimation, and they are generally in line with each other. The Hausman tests for both IV and CF estimation are not significant with p-values equal to 0.307 and 0.219, respectively, and thus, we cannot reject the null hypothesis of exogeneity of those explanatory variables in the model. The result of the Hausman test indicates that the potential endogeneity bias caused by imperfect ability measures and possible reverse causality is not significant.

2.7 Conclusion

In this study, we investigate the effects of an individual's human capital on leadership. We employ more detailed indicators of an individual's human capital (i.e., numeracy, literacy, and problem solving as cognitive abilities; perseverance, openness to learning, and social trust as noncognitive abilities), with the newly available data from the PIAAC. In addition, we adopt various measures of leadership based on job complexity and occupation. We also investigate the potential endogeneity issue caused by imperfect ability measurement and potential reverse causality.

We find that among three cognitive measures, problem-solving ability shows the strongest effect on leadership, and is positive and significant in most cases. A one unit increase in problem-solving test scores will increase the probability of being a leader by 0.273 to 0.3. On the other hand, literacy and numeracy do not show significant effects in general; but in some cases, for example, when leadership is defined in a looser way or in alternative estimation methods, literacy and numeracy also show positive and significant effects. Among noncognitive ability measures, perseverance consistently exhibits a positive and significant effect on being a leader. Individuals with high levels of perseverance have an 8.4 to 9.0 percentage point higher chance of being in a leadership position compared to those with low levels of perseverance.

The commonly used human capital measures, education and experience, have positive and significant effects on being a leader as well. A college degree is expected to increase an individual's probability of being a leader by 7.2 to 8.5 percentage points, and one additional year of work experience increases the probability of being a leader by 0.5 to 0.6 percentage points.

Moreover, the effects of human capital become even stronger when leadership is defined in a more rigorous way, especially for problem-solving ability and education. In addition, the statistic tests and results based on different assumptions about the nature of ability do not show a significant bias caused by the imperfect ability measurement and by the potential reverse causality.

Leadership is vital to the development of any organization. In the US, only 31 to 55 percent large US corporations have a specific mechanism for systematically identifying leadership potentials of their employees (Dries and Pepermans 2012). There is a pressing need for a set of criteria and guidelines that firms can use to assess leadership potential. The development of such criteria is challenging. This study attempts to provide some insights for organizations to develop such mechanisms to identify individuals with strong leadership potentials and facilitate more efficient job assignments and promotions.

There are a few unresolved issues related to this research. The first is about whether the current existing measures can capture one's true inner ability and whether they are correlated with the uncaptured part of the ability. The second issue is to what extent one’s inner ability can be changed because of the environment (e.g., workplace). There has been no consensus among psychological and economic studies on those issues. However, those issues will affect empirical model specification and estimation methodology for any economic studies. Although a further investigation on these questions is beyond the scope of this study, our work should shed new light on them for future studies.

CHAPTER 3

DEMAND FOR MOOC-AN APPLICATION OF BIG DATA

3.1 Introduction

Massive open online course (MOOC) is a tuition-free online courses, in which anyone with internet access can enroll. The New York Time has declared 2012 as "The Year of the MOOC". In 2013, Georgia Tech launched an online Computer Science Master's degree by cooperating with the MOOC provider Udacity. In 2015, Arizona State University (ASU) started to offer MOOC-based academic credits from edX. Up to July, 2015, the number of universities offering MOOCs has increased to 400, and 22 of the top 25 US universities in US News World Report rankings are now offering courses online for free (Shah 2014). The total number of courses in MOOC platform has raised to 2400.

The rapid development of MOOC raise a number of important questions. Will it substitute traditional face-to-face education and bring a revolution to future education? The rising cost of traditional education has been one of the greatest challenges for educators and policy makers. Administrators have adopted several cost-management actions such as increase class size, raise teaching loads, and reduce support staff. However, even with these efforts, the college tuition cost is still viewed as expensive by the general public (Bass 2014). The appearance of MOOC represents a possible solution to the high-cost higher education.

In addition to low cost, MOOC has the advantage of openness and flexibility, which challenges the closed and privileged nature of academic knowledge in traditional education (Krause and Lowe 2014). Working professionals now have better access to education without the limit of geography, time, and financial constraint. In addition, the growth of MOOC offers a plausible solution for the increasing demand for higher education, especially in developing countries. People form less developed areas have the opportunity to study courses from the most prestigious universities, where their academic achievement and future development will be positively shaped by high-quality teacher human capital (Clotfelter et al. 2010). Moreover, MOOC widens the access to high quality education, and lifelong learners have more opportunities to improve their knowledge and skills (Schuwer et al. 2015).

Given the dramatic differences between MOOC and traditional education and the potential of MOOC to our society, it is important to have an in-depth understanding of MOOC. In particular, factors affect individuals' choice of MOOC will have important implications for MOOC's future. Since MOOC is gaining popularity all over the world, we intend to study the demand for MOOC from an international perspective. Specifically, we conduct studies among OECD countries and in China. Most OECD countries are developed countries, while China is one the largest developing countries. The differences in cultural background, economic development, and political system between OECD countries and China provide a relatively complete picture regards the demand of MOOC.

It is challenging to get a precise estimation of MOOC demand, because MOOC platforms (i.e., websites) can be accessed from all over the world. In addition, multiple platforms providing MOOC in different geographic units make it harder to obtain website traffic data based on place of origin. As a result, traditional measures of education demand such as number of college applicants or school enrollment rates cannot be used (i.e., Glewwe and Jacoby 2004). In the current study, we use data from search engine to proxy the demand of MOOC in different geographic units. More specifically, we use search-volume index of MOOC-related keywords generated by search engines. With the rapid development of information technology, number of internet users has accounted an increasing proportion of the world population. The online search behavior of these internet users all over the world have been stored and analyzed by search engines such as Google. It is hypothesized that the more people search for MOOC related terms within a geographic unit, the higher the demand of MOOC will be in that place. The use of search engine data has drawn increasing attention in recent years with the development of information technology, especially the internet. Researchers have used search engine data to forecast various phenomenon such as diseases, unemployment rates, tourist volumes, and housing market (e.g., Ginsberg et al. 2009, Choi and Varian 2012, Yang et al. 2014).

This study adds to the literature by employing big data technology to proxy MOOC demand. The application of such big data allows us to conduct analysis on the demand of MOOC, in which case alternative data sources are very limited. In addition, we attempt to be the first study to investigate different factors that affect MOOC demand and to evaluation this new delivery mechanism of modern education. A comprehensive understanding of such demand can help shape education policy and strategy, such as the investment in internet infrastructure and development of high-demand MOOC courses, for both governments and educational institutes in the era of Internet and the rapid development of information technology.

3.2 About MOOC

The term “MOOC” is first used in 2008 to describe an open online course “Connectivism and Connective Knowledge” (also known as CCK08) designed by educators George Siemens and Stephen Downes. CCK08 was presented to 25 tuition-paying students at the University of Manitoba, and was provided to 2,200 members of general public participates who took the class simultaneously for free with computers and internet. MOOC did not become well known until 2011. In 2011, two professors from Stanford University, Sebastian Thrun and Peter Norvig, designed the first American MOOC, “Introduction to Artificial Intelligence”, which attracted more than 160,000 students. Later, Thrun started up the company Udacity, a platform offering MOOC. Within a year, two other Standard professors, Andrew Ng and Daphne Koller, launched another MOOC platform called Coursera. In 2012, MIT and Harvard launched edX. In recent years, Coursera, Udacity, and edX become the leading MOOC providers and are considered as the “Big Three” MOOC providers (McGuire 2014).[31]

According to a report from Class Central, a MOOC aggregator that collects and posts courses information from various MOOC platforms, Coursera, Udacity, and edX offered more than half of all MOOCs in 2015 (Shah 2015). As in February, 2016, Coursera had more than 1800 courses and more than 18 million users from more than 190 counties. Coursera partners with more than 140 schools all over the world, such as Yale University in the U.S., University of Edinburgh in U.K., and Peking University in China; and edX offers more than 850 courses and partners with more than 90 institutes such as

[31] Coursera website: https://www.coursera.org/, Udacity website: https://www.udacity.com/, edX website: https://www.edx.org/, Big Three MOOC providers article: http://www.nytimes.com/2012/11/04/education/edlife/the-big-three-mooc-providers.html?_r=0

UC Berkeley and The University of Texas System; while Udacity offers more than 120 courses and partners with institutions such as San Jose State University and AT&T. Among three platforms, Udacity concentrates on courses that provide specific skills for workplace, such as computer science, programming and math, while Coursera and edX offer courses in a variety of subjects. Coursera also offers courses in multiple languages such as English, Chinese, French, Spanish, and Italian.

Not only in the U.S., MOOC platforms have appeared all around the world. Famous platforms include Eliademy in Finland, openHP in Germany, FutureLearn in UK, and OpenClassrooms in France. In China, the most significant MOOC platforms include edX, Coursera, as well as local platforms such as XuetangX, IMOOC, and Chinese University MOOC.[32] Specifically, XuetangX is the first Chinese MOOC platform offering courses in all subject from top universities in China such as Tsinghua University and Nankai University. IMOOC is another popular Chinese MOOC platforms offering courses mainly about information technology. While edX and Coursera can provide courses to Chinese from top universities in the world, local platforms better suit their needs because the courses are in Chinese language.

The popularity of MOOC can also be reflected by the fact that an increasing number of education institutions have started to accept MOOC-based credits (Lequerica 2016). For instance, Arizona State University (ASU) and edX announced the Global Freshman Academy in 2015. Students who pass the online courses through edX can obtain college credits from ASU.[33] In addition, it is also possible to get the entire degree

[32] XuetangX website: http://www.xuetangx.com/, IMOOC website: http://www.imooc.com/
Chinese University MOOC: http://www.icourse163.org/
[33] ASU's program: http://asuonline.asu.edu/

based on MOOC. In 2013, Georgia Tech launched an online Computer Science Master's degree by cooperating with Udacity and AT&T.[34] Other online programs include University of Illinois' online MBA degree on Coursera and MIT's (half-MOOC, half-on campus) supply chain management master's degree on edX. [35]

MOOC has attracted an increasing number of students. The total number of student who signed up for at least one MOOC course reached 35 million in 2015, which doubled the number in 2014 (Shah 2015). Although there is no formal study on the demand for MOOC, some quantitative information on MOOC users is available. In particular, regarding why an individual takes MOOC, the most important reasons for taking a MOOC include gaining knowledge and skills, promoting career development, etc. Based on the data from a random end-of-the-course survey in the MIT's first MOOC course (i.e., Circuits and Electronics), Breslow et al. (2013) reported that 55.4% of the surveyed students was for the knowledge and skills they would gain, 25.5% for personal challenge, and 8.8% for employment or job advancement opportunities.

Most MOOC students are highly educated. Christensen et al. (2013) examined data from an online survey of students enrolled in at least one of the University of Pennsylvania's 32 MOOCs offered on the Coursera platform. The result showed that MOOC takers are highly educated, with 83.0% having a post-secondary degree (2 or 4 years), 79.4% having a Bachelor's degree or higher, and 44.2% have a beyond Bachelor's degree, which are much higher than the general educational attainment of their national peers. Additionally, MOOC registrants tend to be young. Grainger (2013) reported

[34] Georgia Tech's program: https://www.omscs.gatech.edu/
[35] University of Illinois's program: https://onlinemba.illinois.edu/, MIT's program: http://micromasters.mit.edu/

analysis of four MOOCs operated by The University of London International Programmes on Coursera in June 2013. The analysis of a pre-course demographic survey showed that male to female ratio is 64:26, and the average age of MOOC users is 34.

The above features about MOOC users provides useful information in studying the demand for MOOC. We will discuss how to measure MOOC demand and specify factors affect its demand in details in later sections.

3.3 Demand for MOOC

A key factor affecting MOOC demand is its relation to traditional college education. Clearly, online education is closely related to traditional education. However, it is less clear whether they are complements or substitutes. On one hand, MOOC's absence of tuition and fees challenges the economic model traditionally subscribed to colleges and universities, especially with the growing cost of college tuition. This new and low-cost form of education is attractive to customers, who can take free and high quality courses online, and meanwhile save money on courses or even an entire degree. With the development of MOOC, students from less developed countries now have access to knowledge from world's leading research institutions. As a result, proponents of MOOC believe MOOCs will bring a revolution to, or even substitute traditional high cost education (Barber et al. 2013). On the other hand, many educators believe MOOC cannot substitute the traditional face to face education. As class size expands, it will be hard for professors to give individual attention and progressive feedbacks in online education (Bass 2014). Additionally, MOOC has a low completion rate, and a significant number of MOOC students stop accessing MOOC courses after the first week (Perna et al., 2013).

Another strand of studies indicate that MOOC can be complement to traditional education because the integration of MOOC into traditional lecture generate promising results with regards to student performance. In the study of the University of Maryland system that incorporated MOOC content into the hybrid courses, students in hybrid courses did as well as or slightly better than students in traditional sections in terms of pass rates and learning assessments (Griffiths et al. 2014).

Therefore, whether MOOC is a substitute or a complement to the traditional education is still up to debate. According to Navarro (2015), the extent to which that online education may substitute traditional classroom delivery depends both on class size and the nature of the institution itself. More specifically, the traditional large hall principle courses will probably be replaced by online education. Large public schools may start to build online courses, while small private institution may advertise “personal classroom” as a differentiator in the education market.

According to the economic principle, demand depends on its own price, the price of related goods, income, preference, and population size. Specifically, opportunity cost of time is part of the own price that affect the cost of MOOC. Since taking classes consumes a considerable amount of time and energy, individuals with a lower opportunity of time (e.g., unemployed individuals) can enroll in either traditional education or MOOCs. Compared to traditional education, MOOC has the advantage of low cost because unemployed individuals tend to have budget limits. On the other hand, for individuals who have higher opportunity cost of time and still want to pursuit an improved knowledge stock (e.g., employed individuals), MOOC becomes a high quality substitution for traditional education because of its flexibility.

Studies have shown positive income effect for traditional education (e.g., Glewwe and Jacoby 2002). According to Maslow's hierarchy of needs, when basic needs such as food and shelter are satisfied, individuals will move up the hierarchy and fulfill higher level of needs such as self-actualization. We expect an increase in income will increase the demand for education, which helps to achieve higher level of self-fulfillment. In addition, higher income promotes online education because high income individuals are more active in technology use. Studies indicated that low-income individuals face greater challenges when using such online resources because they have lower perceived technical skills (DiSalvo et al. 2016).

Preference towards MOOC can be affected by factors such as education level, or internet speed. According to Moon (2004), high knowledge level individuals are more likely to conduct online information search because they have higher search proficiencies. In addition, they may prefer to enroll in online education because they are more motivated to learn new knowledge and open to new technology. In addition, internet speed can change individuals' preference/interests. A slow internet connection increases the opportunity cost of time and reduces students' interests in learning.

To investigate the demand for MOOC, it is critical to find a measure for such demand. Differing from traditional education where demand can be measured by number of applicants or enrollment rates, it is challenging to get a good estimation of MOOC demand, mainly because MOOC is online, and applicants can be from all over the world. In addition, private registration data for various MOOC platforms are not available for public use.[36] As a solution, we attempt to proxy MOOC demand from the perspective of consumer information search and purchase behavior through internet.

With the rapid development of Internet, an individual's demand can be reflected by Internet search (Peterson and Merino 2003). According to McGaughey and Mason (1998), internet influences buyer behavior through each step of the classical buyer decision process, including problem recognition, information search, evaluation of alternatives, purchase decision, and post-purchase behavior. In particular, individuals realize the need for additional education/courses (i.e., problem recognition) and then gather related information (i.e. information search). During this search process, they use search engines and search keywords that can lead them to their specific areas of interest. For example, for individuals who want an improved knowledge and decides to choose online education, they may search more general keywords such as "(free) online education" or "(free) online course" at the beginning and then realize the existence of MOOC. Then an in-depth search of MOOC will expose individuals to various platforms such as Coursera, Udacity, and edX (i.e., alternatives). Later on, they just need to search the name of certain platforms to find more details.

As a rational individual, the evaluation process is a comparison of utility brought by different alternatives. Particularly for different MOOC platforms, Coursera and edX have a wide range of courses choice in all subjects, while Udacity concentrates more on job market skills. Based on personal characteristics such as financial constraint, time limit, and preference, they will choose the product that best satisfy their needs. The above

[36] There are companies evaluating web traffic data such as Semrush and Alexa. However, the information they provided are largely constrained by time and location. For instance, in countries where Google is not commonly used (i.e., China), Semrush cannot provide estimated traffic data.

Internet search behaviors are captured, stored, and analyzed by the search engine. Using such information can give us an estimation of how many individuals in a certain area have searched certain keywords.

With a proper choice of keywords, search volumes can properly reflect the potential MOOC demand. Specifically, search volume for a more general term such as "MOOC" may only reflect how many people are interested or just curious about what MOOC is. However, search volume for certain platforms such as "Coursera" can capture individuals who are seriously interested in MOOC and highly likely to register. Therefore, with a combination of general and specific keywords, we use search index generated by search engines to represent the demand for MOOC. In our study, we use keywords including "MOOC", "Coursera", "Udacity", and "edX".

Based on the above procedure, we collect two sets of data, one is a cross-country dataset among OECD countries and the other a within country dataset for China. With these data, we estimate the demand for MOOC. OECD includes mostly developed countries, while China is a developing country but with significant economic gap between coastal and inland provinces. China and OECD countries are very different in cultural background, economic development, political system, and language. Such different data sets can provide relatively complete picture regarding the demand for MOOC, and thus help understanding the future of MOOC.

3.4 Demand for MOOC in OECD countries

Registrants from OECD countries account for a significant portion of worldwide MOOC students. Based on a survey by Christensen et al. (2013), more than 65% of surveyed MOOC students originates from OECD, with 34.3% coming from the United States, and among the rest of the registrants, 14.8% originates from BRICS (i.e., Brazil, Russia, India, China, and South Africa). In addition, according to an article published by Coursera (2016), four of its Top 10 countries of Coursera learners are OECD countries including United States, Canada, United Kingdom, and Spain.[37] MOOC is also getting political support in OECD countries. For instance, in recent years, the European Commission funded a number of MOOC projects such as HOME project-Higher Education Online: MOOCs the European Way to develop and strengthen an open network for European cooperation, especially for MOOC.

To study the demand for MOOC in OECD countries, we use Google Trends data to proxy MOOC demand in each country.[38] Google Trends provides (nearly) real-time Google query data from January 2004 to present on a weekly or monthly basis. It does not report the raw search volume; rather, it reports a query index, which displays total searches for the keywords in relative to the total number of searches done on Google over time given the specific location and time range. Then, all the search indexes are divided by the highest "relative search rate" within a search. As a result, Google Trends search indexes have a range of 0 to 100. Since an increasing trend in Google Trends could either result from more absolute search for that keyword or fewer other searches, special values of Google Trends cannot be specifically interpreted. To deal with this issue, we convert the Google Trends index into an absolute search volume index based on the following formula:

$$GT_{it} = KSV_{it}/TSV_{it}\,/a$$

[37] The link to the Coursera' blog https://blog.coursera.org/post/142363925112
[38] Google Trends can be accessed from https://www.google.com/trends/

where GT_{it} is reported Google Trends search index, KSV_{it} is keyword search volume, TSV_{it} is total search volume in country i and year t, and a is the highest "relative search rate". In addition, we use the formula to proxy TSV_{it}:

$$TSV_{it} \approx WSV_t * (IU_{it}/WIU_t)$$

where WSV_t is world total search volume at time t, IU_{it} represents number of internet users in country i and time t, and $WIUt$ is number of internet users in the world at time t. As a result, the absolute search volume index KSV_{it} can be calculated as:

$$KSV_{it} = a * GT_{it} * TSV_{it} = a * GT_{it} * WSV_{it} * \left(\frac{IU_{it}}{WIU_t}\right)$$

We use KSV_{it} index to proxy the demand of MOOC.[39] We expect that the higher the search volume index for MOOC related keywords, the higher the demand for MOOC.

Figure 3 is a snapshot of Google Trends showing the relative popularity of these five keywords. Consistent with reality that MOOC took off in 2012, Figures 3 also shows that the popularity of MOOC and its platforms surged from 2012. The high points are usually in January to February, or September to October, which corresponds to the start of school spring or fall semester. The low points are usually in December, which corresponds to the holiday season. In addition, the Figure 3 shows that the search for "Coursera" is much higher than "Udacity", "edX", and "MOOC". The possible reason is that after individuals get exposed to the platforms of MOOC, they will probably search the name of the platform directly next time. As a result, the most popular platform (i.e. Coursera) will have more search volumes. Moreover, Figure 3 shows that the search volume index for the combination of four terms is the highest. Considering the combined

[39] KSV_{it} can be viewed as a search index that represents absolute search volume. It contains an unknown constant a.

search term is the most general one, we use this combination to proxy MOOC demand in each OECD country. We first obtain monthly index of MOOC demand for each country from Jan 2012 to December 2015 and then calculate the yearly average.

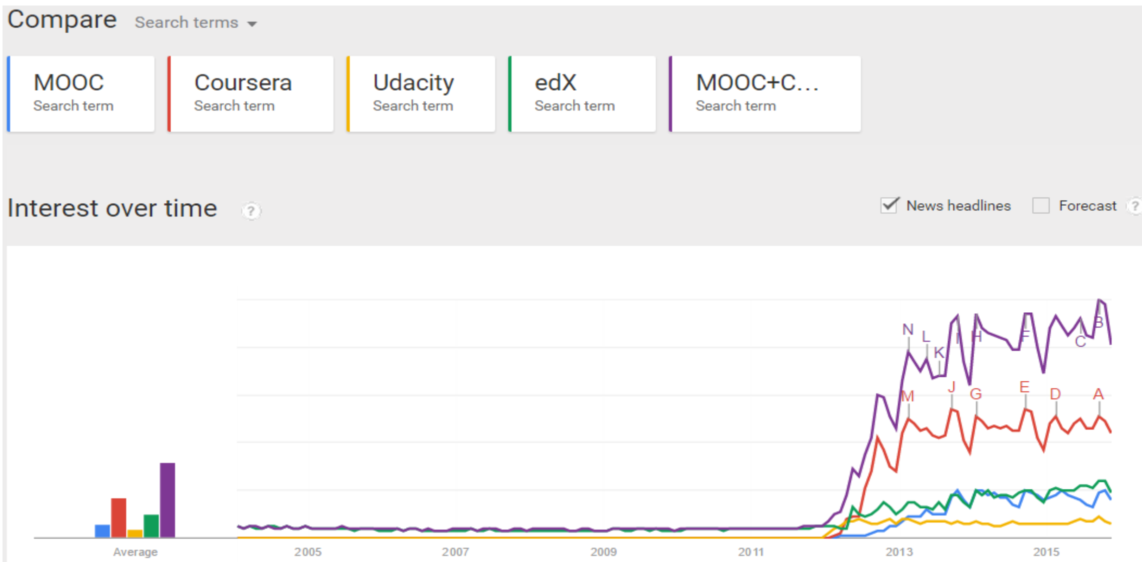

Figure 3. A snapshot of Google Trends search

Table 8 lists the descriptive statistics.[40] For comparison purpose, we list the first (i.e., 2012) and the last year (i.e., 2015) of our data. Table 8 shows that MOOC demand measured by the KSV index increases significantly from 2012 to 2015. The demand in 2015 is approximately 2.6 times higher than that in 2012. To see the cross country differences of KSV index, Figure 4 displays the index of each OECD country in 2015. It shows that United Kingdom, Spain, Mexico, France, and United States are countries with highest volume, while Iceland, Estonia, Luxembourg, Slovenia, and Slovak Republic are among countries with lowest search volume of MOOC. Since KSV index indicates absolute search volumes, higher/lower indices are normally associated with higher/lower number of population.

[40] Percent of college and high school graduates for 2014 and 2015 are linearly predicted based on data from 2010 to 2013. Wage data for 2015 are linearly predicted based on historical data from 2000 to 2014. Internet speed data for 2015 are linearly predicted based on historical data from 2012 to 2014.

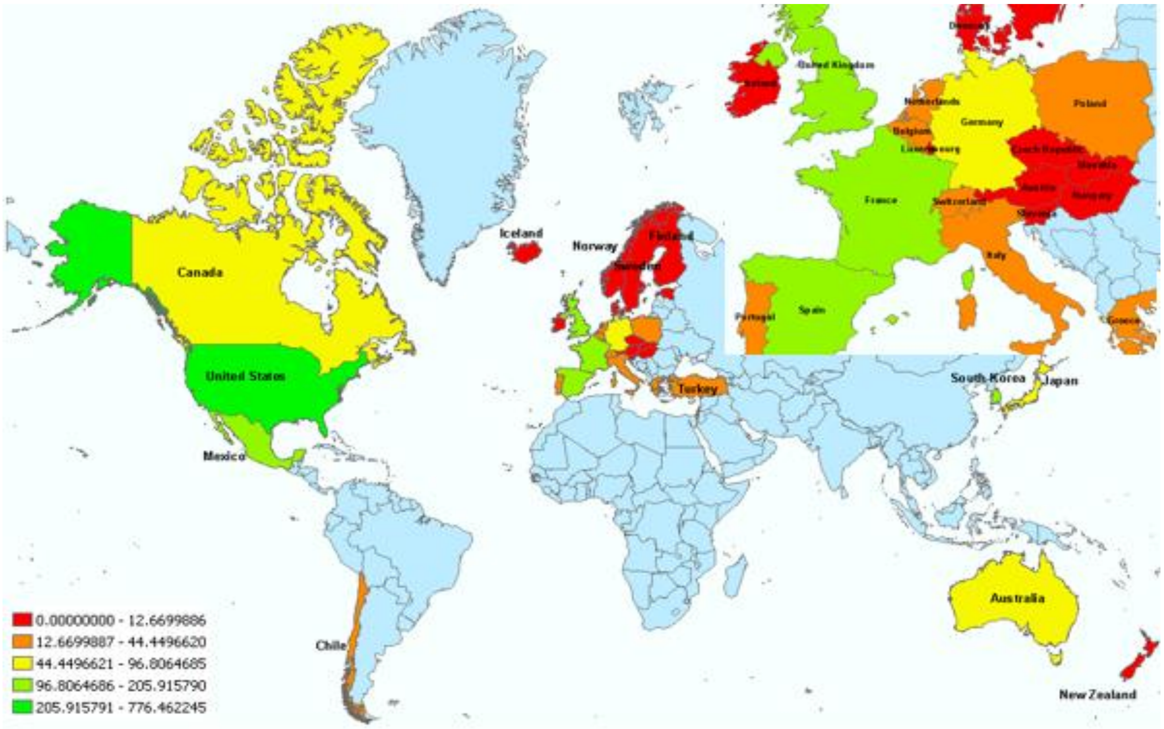

Figure 4. KSV Index for OECD countries in 2015

As discussed in the theoretical model, variables affect MOOC demand in OECD countries include unemployment rate, average annual wage, internet speed, education level, and population size.[41] Unemployment rates data are obtained from Labor Market Statistics.[42] Based on this dataset, our unemployment rate variable is calculated as numbers of unemployed people as a percentage of the labor force, which is defined as the total number of unemployed people plus those in civilian employment. Table 8 shows that unemployment rate decreases slightly from 8.71% to 7.97% from 2012 to 2015. In addition, average wage of each country are obtained from OECD Employment and Labor Market Statistics and are measured in constant prices at 2014 US dollars.[43] Table 8 shows that average annual wage increases by approximately 1000 US dollars during the period.

[41] College tuition is not included because of data limit in OECD countries.
[42] Data source: https://data.oecd.org/unemp/harmonised-unemployment-rate-hur.htm
[43] Data source: https://data.oecd.org/earnwage/average-wages.htm

Table 8. Descriptive Statistics for OECD countries

	Year 2012				Year 2015			
Variable	Mean	S.D.	Min	Max	Mean	S.D.	Min	Max
KSV	26.33	70.78	1.06	406.54	65.87	137.13	3.57	776.46
Unemployment	8.71	5.04	3.18	24.81	7.97	4.76	3.40	25.47
Wage	38.24	12.80	12.71	58.33	39.74	13.11	12.54	61.53
Internet speed	14.35	6.72	4.77	32.05	28.63	12.70	8.66	58.42
Percent of college graduates	32.57	10.03	15.30	52.60	34.91	10.46	17.07	55.10
Percent of high school graduates	43.88	14.11	18.57	73.17	43.67	13.21	18.77	71.99
Population	12.70	20.10	0.11	105.51	12.67	20.30	0.11	106.91

Note: 1. KSV index represents absolute search volumes of MOOC, which is calculated based on Google Trends.
2. Unemployment rate, wage, education level, and population data are obtained from relevant OECD databases described in the data section. Wage is measured by thousand dollars in constant price at 2014 US dollars. Education level is represented by percent of college and high school graduates between the age group of 25-64 in OECD countries. Population is measured by millions of population between 20 to 44 years old.
3. Internet speed is measured by download speed in Mbit/s and the data are collected by Ookla.
4. Due to data limit, percent of college and high school graduates for 2014 and 2015 are linearly predicted based on data from 2010 to 2013. Wage data for 2015 are linearly predicted based on historical data from 2000 to 2014. Internet speed data in 2015 are linearly predicted based on data from 2012-2014.

Internet speed are obtained from a series of OECD reports (2014, 2015). It is measured by download speed (Mbit/s) of the first quarter of each year collected by Ookla, a global leader in broadband testing and web-based network diagnostic applications. Based on Table 8, internet speed increases significantly during the period, from 14.35 Mbit/s in 2012 to 28.63 Mbit/s in 2015. Specifically in year 2015, countries with highest internet speed include Korea, Japan, Switzerland, Sweden, and Netherland, and countries with lowest internet speed include Italy, Greece, Turkey, Mexico, and Australia. United States only has intermediate internet speed due to the less of competition (Miller 2014).

Percentages of college and high school graduates between the age group of 25-64 are obtained from Education dataset: Population who attained tertiary education by sex and age group.[44] Percent of college graduates is represented by the percent of tertiary education, which is equivalent to ISCED 5 or 6 based on the 1997 International Standard Classification of Education (ISCED 1997). In addition, percent of high school graduates is represented by the percent of upper secondary education, which corresponds to ISCED 3. According to ISCED 1997, ISCED 3 level typically begins at the end of full-time compulsory education for those countries have a system of compulsory education and the entrance age to this level is typically 15 or 16 years. Table 8 shows that the percent of college graduates increases approximately 2% from 2012 to 2015 and the percent of high school graduates remain relatively stable. [45]

To measure the size effect, we use population data from 20 to 44 years old, which is also the range of population that are most likely to enroll in MOOC courses. Population data are obtained from OECD.Stat.[46] Table 8 shows that population between 20 and 44 years old stays relatively stable from 2012 to 2015.

Table 9 displays estimation results for OECD countries. Model 1 presents the fixed effect estimation, in which country fixed effects that do not change with time are removed.[47] Model 2 controls for year dummies in case unexpected variation or special events in a particular year that could affect MOOC demand. Model 3 controls for a linear

[44] Data source: https://data.oecd.org/eduatt/adult-education-level.htm

[45] In 2012, the maximum percent of high school graduate is 73.17% in Czech Republic, which is lower than the maximum percent in 2015 (i.e., 71.99% in Slovak Republic). In 2015, the percent of high school graduates in Czech Republic drops to 71.73%.

[46] Data source: http://stats.oecd.org/Index.aspx?DatasetCode=POP_FIVE_HIST

[47] We also run the basic OLS estimation and the Random Effect model. The results are consistent with fixed effect estimation except for that percent of high school education becomes negatively significant. However, the Hausman test rejects the Random Effect model.

common trend for each country to capture the overall movement of MOOC demand across years. In addition to a linear trend, Model 4 also controls for the square term of the trend in case the increase in MOOC demand is nonlinear.

Model 1 indicates that higher unemployment rate will increase MOOC demand. Specifically, if unemployment rate increases by one percent, MOOC demand (i.e. KSV index) increases by 0.144%. It is consistent with our prediction in the theoretical model because unemployed individuals have lower opportunity cost of time. According to a report by Krueger (2008), unemployed American who spent some time looking for work devote an average of two hours and 40 minutes in job search activities, compared with the average of one hour and 40 minutes in 13 other OECD countries. This leaves most unemployed individuals with considerable time to improve their skills and get a new job. In addition, unemployed individuals have stronger motivation to take MOOCs to expand their knowledge and search for a new job. This is consistent with previous findings about MOOC, which showed that the most important reasons for taking MOOC was for knowledge and skills, interests, and employment or job advancement (e.g. Breslow et al. 2013).

Table 9. Factors affect MOOC demand in OECD countries

	FE	Dummy	Trend	TrendSQ
Unemployment	0.144***	0.076***	0.129***	0.092***
Wage	0.549	0.501	-3.589	0.878
Internet speed	0.766***	0.1	-0.016	0.001
Percent of college graduates	0.207***	0.079**	0.088	0.086**
Percent of high school graduates	0.106*	0.082**	0.065	0.083**
Population	-1.103	0.749	-0.197	0.488
Year Dummies 2013		0.926***		
Year Dummies 2014		0.982***		
Year Dummies 2015		0.917***		
Trend			0.361***	1.541***
Trend2				-0.249***
Constants	-12.424	-8.116	8.3	-10.414
N	119	119	119	119
R-squared	0.598	0.896	0.653	0.869
F	20.593	76.774	22.037	67.304

Note: 1. Dependent variable is KSV.
2. Due to data limit, OECD study does not include tuition information. We use variable "percent of college graduate" to discuss the complement or substitute effect between MOOC and traditional college education.
3. *** represents significance at 1% significant level, ** represents significance at 5% significant level, and * represents significance at 10% significance level.

The significant impact from unemployment on MOOC demand also indicates that MOOC may help reduce unemployment. Coursera CEO Richard Levin, who used to serve as the president of Yale University for 20 years, notes that Coursera and other MOOC providers are going to potentially have a major impact on solving some of the structural unemployment concerns both in the U.S. and abroad by providing necessary and in-demand skills in job market (Chui, 2015). For instance, Udaicy has several hiring partners such as Google and AT&T that are regularly hiring their talented graduates. By providing students extensive and personalized career support, graduates from Nanodegree Plus program provided by Udacity are guaranteed to get a job within six month of graduation date; otherwise, they will get 100% of their tuition refunded. In addition, Flipkart, an e-commerce company headquartered in India, starts to hire students based on Udacity's Nanodegree programs without interviews, which indicates that MOOC courses could potentially increase a lot of working opportunities (Ayyar 2016).

As discussed in the theoretical model, higher wage is expected to increase MOOC demand by substitution and income effect. For substitution effect, MOOC can be a high quality substitution for traditional education due to its flexibility, especially for individuals with high wages and have a high opportunity cost of time. However, our results indicate that the impact of wage on MOOC demand is not significant. The possible reason is that the substitution and income effect brought by increased wage is not strong enough in OECD countries.

In addition, Model 1 shows that internet speed has a positive impact on MOOC demand and it is statistically significant. Higher internet speed enables a fast and reliable access to MOOC, which will reduce the opportunity cost of time and increase the preference/interests for MOOC. Specifically, a one percent increase in internet speed will increase MOOC demand by 0.766 percent. Higher internet speed has been considered as an important factor in teaching and learning success. Federal Communications Commission (FCC), provides a Broadband speed guide to estimate minimum internet speed required for certain online activities.[48] According to the report, HD-Quality streaming movie or university lecture requires a minimum of download speed at 4 Mbps/s.

Moreover, Model 1 shows that a one percent increase of individuals with a college degree increases MOOC demand (i.e. KSV index) by 0.207%. This positive

[48] The guide can be accessed from https://www.fcc.gov/reports-research/guides/broadband-speed-guide

impact is also supported by the literature showing that most MOOC takers already have a Bachelors' or above degree. As discussed fore, it is possible that these highly educated individual have stronger motivation to learn new knowledge and are more familiarly with online technologies. In addition, we find that an increase in the percentage of high school education also promotes MOOC demand, with a smaller elasticity of 0.106, which indicates that individuals with high school education also choose MOOC as a knowledge source. Population size do not show significant impact on MOOC demand.

Model 2 display the fixed estimation result with time dummies. All three year dummies are positive and significant, indicating the existence of year-specific fixed effects that are not controlled by other explanatory variables. In addition, unemployment and education level variables are still significant, but with smaller magnitudes. The possible reason is that part of their impacts are absorbed by time dummies. In addition, internet speed becomes in significant after adding time dummy, indicating that the impacts from internet speed are captured by time dummies.

Models 3 and 4 report the estimation result with linear trend and both linear and nonlinear trends. According to Model 4, both trend and trend^2 are significant, indicating that Model 4 is preferred than Model 3. In addition, since trend is positive and trend^2 is negative, MOOC demand has a general upward trend and the demand increases faster at the beginning and then slows down, which is consistent with Figure 3. The rest of the variable estimations are very similar to Model 2.

In this section, we investigate the demand for MOOC using cross country data. However, due to drastic differences in market structure, social institutions, and education system across countries, a single unified demand function may not work well. In addition, most OECD countries are developed countries, so the analysis result may not be generalized to developing countries. As a result, we conduct another analysis about the demand for MOOC within a single country, China, which is one of the largest developing country and administers the same education system and policy structure.

3.5 Demand for MOOC in China

With the largest higher education market and the largest number of internet users in the world, MOOC has a huge potential in China. In July 2015, Coursera announced that they have more than 1 million registrations from China, making the country their second largest after the United States (Shah 2015). MOOC has also gained political support from Chinese government. According to a document about online courses issued by the Ministry of Education of the People's Republic of China (MOE) in 2015, government, college, and society should work together to promote the development of online education platforms. By the year of 2020, China should have at least 3000 nationally recognized high quality online courses.[49]

We use Baidu Index to proxy the demand for MOOC in each province in China.[50] Baidu Index is a similar service with Google Trends, which provides Baidu query volume data from June 2006 to the present on a daily basis. Baidu search index reflects absolute numbers and are directly comparable. For comparison purposes, we use the same keywords as with the OECD study, including MOOC, Coursera, Udacity, and edX. Figure 5 shows the snapshots of Baidu Index of four keywords. Based on Figure 5, in China, MOOC and its three most popular platforms started to gain search volume from

[49] The link to the document: http://www.moe.edu.cn/publicfiles/business/htmlfiles/moe/s7056/201504/186490.html

[50] Baidu index can be accessed from http://index.baidu.com/

2012, developed steadily in 2013, and became relatively stable in year 2014 and 2015. The high points are oftentimes associated with the release of news or journal reports containing keywords of MOOC and the low pints always fall in the week of Spring Festival, the most traditional and important holiday in China.

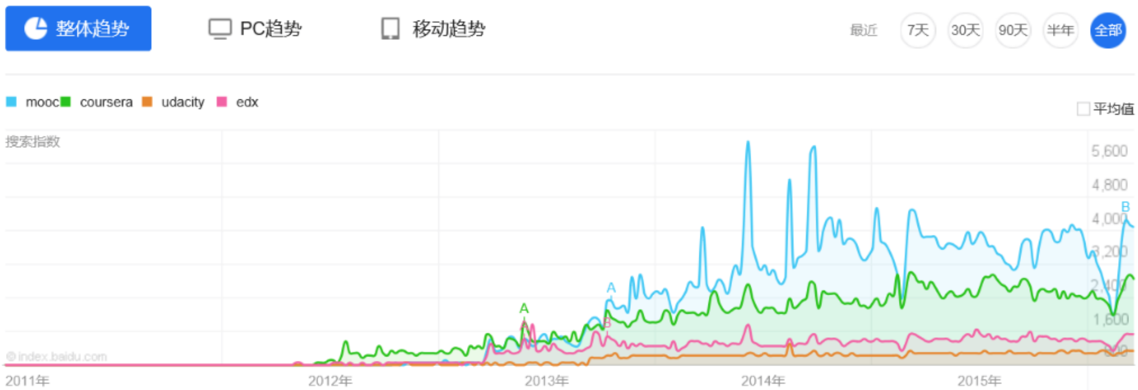

Figure 5. A snapshot of Baidu Index search

Table 10 presents the descriptive statistics.[51] Baidu index increases significantly since their appearance. In specific, the search in 2015 is 26 times higher than 2012. To better illustrate the cross-province difference of MOOC demand in China, Figure 6 presents the Baidu search volume data in each province. It is clear that Beijing and Guangdong province have the highest interests in MOOC. In addition, coastal provinces (i.e., east side) such as Jiangsu and Zhejiang have higher Baidu index compared to inland provinces (i.e., west side) such as Xizang and Xinjiang.

[51] Due to data limit, tuition per capital, percent of college and high school education, unemployment rate, GDP per capita, and internet users in 2015 are linear predictions based on available historical data. Internet speed data for 2012 are predicted based on 2013-2015 data.

Table 10. Descriptive Statistics for China

	Year 2012				Year 2015			
Variable	Mean	S.D.	Min	Max	Mean	S.D.	Min	Max
Baidu	0.18	0.11	0.00	0.47	4.76	2.71	0.31	13.02
Tuition per capita	7.27	2.72	2.60	16.33	7.81	3.45	2.76	19.13
Unemployment	3.32	0.64	1.27	4.23	3.12	0.72	1.21	4.44
Wage	41.05	10.63	31.30	75.59	55.22	13.86	42.18	102.27
Internet speed	0.98	0.55	0.09	2.83	6.01	0.50	5.01	7.10
Percent of college graduates	14.14	7.41	5.32	43.37	16.36	8.62	1.65	48.83
Percent of high school graduates	19.70	4.33	6.40	26.70	21.04	5.07	4.31	28.32
Internet users	18.19	13.41	1.01	66.27	22.40	15.79	1.35	76.31

Note: 1. Baidu represents Baidu search index.
2. Tuition per capita data measured in thousand RMB are calculated based on data from China Education Finance Statistical Yearbook (2011, 2012, 2014) and China Statistical Yearbook (2012, 2013, 2014, 2015).
3. Unemployment, wage, education level data are obtained from China Statistical Yearbook (2012, 2013, 2014, 2015). Wage represents real wage with 2012 as the based year and it is measured by thousand RMB.
4. Internet speed measured by average connection speed in Mbit/s are obtained from China Internet Speed Report from Broadband Development Alliance. http://www.chinabda.cn/xzzq/index.shtml
5. Internet users measured in millions are obtained from Statistical Report of internet Development in China by China Internet Network Information Center (CNNIC). http://www1.cnnic.cn/IDR/ReportDownloads/

Similar to the OECD study, variables affect the demand of MOOC in China include tuition per capita, unemployment rate, average annual wage, internet speed, education level, and number of internet users. Tuition per capita data are calculated by the ratio of government income from tuition and fees and number of enrolled students in college and universities in China. Government income from tuition and fees is obtained from China Education Finance Statistical Yearbook Table 3-7 (2011, 2012, 2014) and number of enrolled students is obtained from China Statistical Yearbook (2012, 2013, 2014, 2015).[52] Table 10 shows that tuition per capita increases from 7300 RMB per year per student in 2012 to 7800 RMB in 2015.[53]

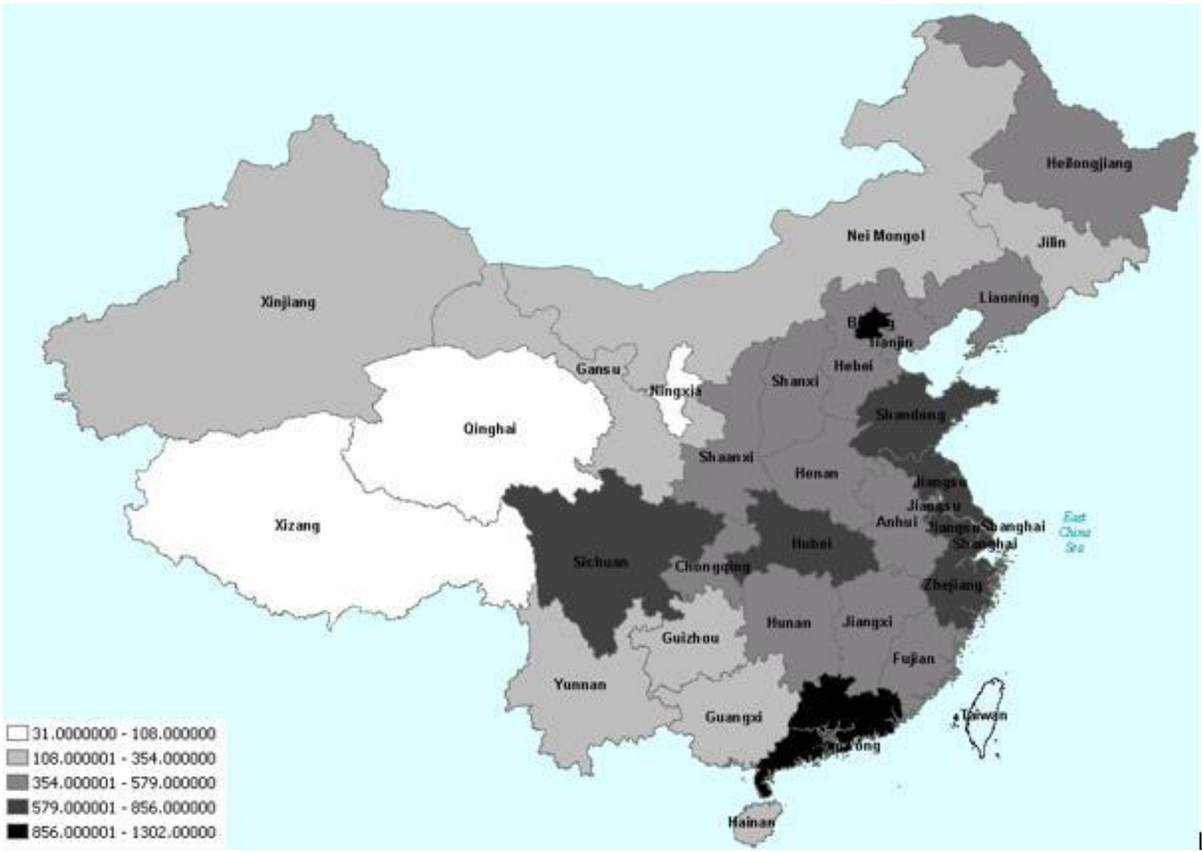

Figure 6. Baidu Index for provinces in China in 2015

Unemployment rate data are calculated as the number of unemployed people in urban area as the percentage of the total number of unemployed people plus those in civilian employment in urban area and it is obtained from China Statistical Yearbook (2012, 2013, 2014. 2015). Table 10 indicates that unemployment rate remains relatively stable in China from 2012 to 2015, with a slight decrease. Average wage data are calculated by total wage divided by the number of employees. We then convert it to real

[52] Due to data limit, government income from tuition and fees in 2013 is the average of data from 2012 and 2014.
[53] The minimum tuition per capital in both year is in Xizang, one of the least developed province in China, and the maximum is in Beijing, one of the most developed city in China.

average wage based on the consumer price index (CPI) with the base year 2012. Both wage data and the CPI are obtained from China Statistical Yearbook (2012, 2013, 2014, 2015). Table 10 indicates that average real wage increases from 41 thousand RMB in 2012 to 55 thousand RMB in 2015.

Internet speed are average connection speed obtained from China Internet Speed Report from Broadband Development Alliance.[54] Table 10 also shows that internet speed in China increase from 0.98 Mbit/s in 2012 to 6.01 Mbit/s in 2015, representing a 6 times increase. Specifically in 2015, provinces/city with highest internet speed include Beijing, Shanghai, and Tianjin, and provinces with lowest internet speed include Xiang, Guangdong, and Gansu.

Percent of population with college education is the ratio of number of college graduates and number of labor force from 16 to 64 years old. Similarly, percent of high school graduate is the ratio of number of high school graduates and number of labor force from 16 to 64 years old. Number of college graduates, high school graduates, and labor force data are obtained from China Statistical Yearbook (2012, 2013, 2014, 2015). The percent of college graduates increase from 14.14% to 16.36%, and high school graduates increases slightly from 19.70% to 21.04% from 2012 to 2015.[55] Internet users are obtained from Statistical Report of internet Development in China by China Internet Network Information Center (CNNIC).[56] Table 10 indicates that the number of internet users increase by approximately 4 million during the period.

---

[54] Data source: http://www.chinabda.cn/xzzq/index.shtml
[55] Note that the maximum percent of college graduates is much higher than that of high school graduates in both 2012 and 2015. The reason is that in Beijing, the percent of college graduates is much higher than high school graduates in both years. For instance, in 2015, the percent of college graduates in Beijing is 49% and high school graduates is 24%.
[56] Data source: http://www1.cnnic.cn/IDR/ReportDownloads/

Table 11 displays the result for MOOC demand in China. Similar to the OECD study, Model 1 reports the fixed effect estimation, and Model 2 and 3 control for year dummies and trend. [57] Model 1 shows that tuition per capita is not significant. Thus, even though the rising cost of college tuition could potentially reduce college enrollment in China, we do not observe any complement or substitute effect from MOOC. The possible reason is that MOOC's impact in China is still limited. MOOC is still developing in China, so it may not be familiar to a lot people, especially in rural and poorer regions. At the current stage, there is no university in China offers MOOC-based academic credits, so it is unrealistic for Chinese students to substitute a college degree with a MOOC based degree.

Model 1 also shows that increased unemployment rate will increase MOOC demand in China. Specifically, a one percent increase in unemployment rate will increase MOOC demand by 1.088%. As we discussed before, MOOC platforms provide job opportunities by providing job specific skills and benefit unemployed individuals, who also have a lower opportunity of time. In addition, we find a higher internet speed will significantly promote MOOC demand. China's internet speed is one of the slowest around the world (Akami 2015). Such a slow connection prevents many Chinese internet users from enrolling in MOOCs that requires a fast and stable connection. As a result, a faster internet speed is necessary in increasing MOOC demand in China.

[57] We also run the basic OLS estimation and the Random Effect model. The results are consistent with fixed effect estimation. However, the Hausman test rejects the Random Effect model.

Table 11. Factors affect MOOC demand in China

	FE	Dummy	Trend
Tuition per capita	1.191	-0.096	1.298
Unemployment	1.088***	-0.156	1.078***
Wage	6.603***	0.163	7.123***
Internet speed	1.181***	0.054	1.178***
Percent of college graduates	-0.065	-0.009	-0.065
Percent of high school graduates	-0.071	-0.018	-0.069
Internet users	-0.348	1.353	0.001
Year dummy 2012		2.407***	
Year dummy 2013		3.073***	
Year dummy 2014		2.986***	
Trend			-0.068
Constants	-21.453	-3.94	-23.981
N	123	123	123
R-squared	0.831	0.985	0.829
F	90.938	822.48	78.716

Note: 1. $Trend^2$ is not included because the linear trend is not significant.
2. *** represents significance at 1% significant level, ** represents significance at 5% significant level, and * represents significance at 10% significance level.

Emerging from Model 1 is a strong and positive impact from wage. If wage increases by one percent, the demand for MOOC (i.e., Baidu Index) will increase by 6.603 percent. The substitution effect and income effect brought by a wage increase reinforce each other, and thus the magnitudes of the impact form wage is relatively large in China. On one hand, the development of MOOC provides a good opportunity for employed workers in china who want to get extra knowledge, and MOOC becomes a high quality substitution for traditional education because of its flexibility and low cost. On the other hand, a higher income may bring adjustments of individual consumption bundles and increase the demand for education related products. Last but not least, we do not observe significant impact on MOOC from increased percentages of college or high school graduates.

After controlling for year dummies in Model 2, all the explanatory dummies become insignificant. One possible reason is that year dummies are highly correlated with our explanatory variables such as internet speed and internet users. It captures a lot of impacts from other explanatory variables and also consumes additional degree of freedom. Model 3 controls for the trend. The linear trend is not significant and the results are in general consistent with Model 1.[58]

Comparing the results from China with OECD countries, we find that both unemployment rate and internet speed positively affect MOOC demand, which is consistent with our theoretical prediction. Different from OECD countries, wage has a positive and significant impact on MOOC demand in China. One possible reason is that the gap between high- and low-income individuals with regards to technology/computer use is more significant in China compared to OECD countries. Many low-income Chinese, especially in rural area, lacks basic skills in using online resources, and a higher income brings a very significant impact in China. According to a survey by Emanuel (2013), almost 80% of MOOC students come from the wealthiest and most well-educated 6% of the population in BRICS countries.

In addition, our results show that an increased percentage of college and high school graduates significantly increase MOOC demand in OECD countries, which may indicate that they choose MOOC as substitutes for traditional college when they need further education. However, we do not find such significant impact from education variables in China. Moreover, an increase in college tuition does not have a significant impact on MOCC demand in China. The possible reason is that college education in

[58] $Trend^2$ is not included because the linear trend is not significant.

China is still scared, so Chinese still prefer traditional education. The results in China also indicate that the substitution effect of MOOC is limited.

In addition to the above international platforms analysis, we also run an analysis based on local Chinese platforms. Chinese local MOOC platforms include XuetangX, IMOOC, and Chinese University MOOC. XuetangX is the first Chinese MOOC platform offering courses from top universities in China and it is operated by Tsinghua University. IMOOC is currently the most popular Chinese MOOC platforms offering courses mainly about information technology. However, these local platforms appear one or two years after international platforms and the most popular platform, IMOOC, appears in year 2014, which leaves us very limited data. Still, we obtain the search volume for local MOOC platforms and run the same analysis with Table 11. The results are presented in Appendix Table 12. Similar to international platform, increase in wage and unemployment also increase MOOC demand for local platforms. However, we find an unexpected and negatively significant impact from internet speed when trend is controlled for.

3.6 Conclusions

Facing various challenges in traditional education, many educators consider MOOC as a possible solution and believe it will bring a revolution to traditional education. In this study, we aim at investigating factors affect the demand for MOOC. We use search engine data in both OECD countries and in China to proxy the MOOC demand and estimate a simplified demand function based on data availability. Our estimation result shows that in both OECD countries and China, increased unemployment rate and higher internet speed will promote the demand for MOOC. In addition, higher education level significantly increase MOOC demand in OECD countries, while wage has a positive and significant impact on MOOC demand in China.

The development of MOOC brings many opportunities to the society. As we mentioned before, MOOC has the potential to reduce the cost of higher education. With the development of information technology, online classes will be more well-presented, flexible, and interactive. Thus, educational institutions should consider build online courses category based on their strengths (Navarro 2015). In addition, MOOC can reduce inequality of education resource distribution by delivering courses from the best universities to all around the world. Governments, especially in less developed areas, should seize the opportunity and take actions to encourage MOOC development. Such actions include investment in internet infrastructure to guarantee a stable connection, workshops about how to search and register for MOOC courses, and government-endorsed certificates to individuals who successfully complete the course. Moreover, MOOC provides a valuable option for lifelong learners, who may have graduated from college but need future education to keep up with the technology development. From this aspect, government should encourage MOOC platforms that train on-demand skills.

CHAPTER 4

FAMILY, TEACHER, AND STUDENT ABILITY DEVELOPMENT IN CHINA

4.1 Introduction

The importance of cognitive ability on an individual's school attainment, career development, and social behaviors has been widely recognized (i.e., Herrnstein and Murray 1994, Heckman et al. 2006). In recent years, noncognitive abilities, which represent personalities, behaviors, and attitudes such as perseverance and self-discipline, are also gaining increasing attention in literature. A series of studies showed that noncognitive abilities are as important as, if not more important than cognitive abilities in determining academic and career outcomes (Heckman and Rubinstein 2001). As a result, understanding the determinants of ability leads to a better understanding of individuals' long run outcomes. The object of this study is to evaluate the impact of family and teacher certifications on the development of cognitive and noncognitive ability among Chinese middle school students.

Previous studies show that family inputs such as family size, financial condition, and parental education are significant predictors of student academic achievement (Hanushek 1992, Rosenzweig and Zhang 2009). Similar, teacher credentials are also considered as the most significant institutional determinant of student achievement (Rockoff 2004, Kane et al. 2006, Clotfelter et al. 2010). However, most of these existing studies measure student achievement by educational outcome or by test scores in specific subjects such as Math, and Science. Much less research has been done about the impact of family inputs or teacher credentials on individual ability.

Different from test scores in a particular subject, which is almost entirely based on the knowledge from schooling, ability measures an individual's intelligence (i.e., cognitive) or personality (i.e., noncognitive), and it is separated from knowledge. Based on physiological studies, cognitive ability develops early in life and reach stability in adolescence or early adulthood at latest (e.g., Deary et al. 2000, Schalke et al. 2013). In addition, it is believed that individuals' cognitive ability becomes stable at a relatively early stage of life (Cunha and Heckman 2007). On the other hand, even though studies also showed stability of different aspects of personality, less consensus exists on the development of noncognitive ability, with some arguing that it can be altered at the end of teenage years and others claiming that it can be changed at any stage of life (Brunello and Schlotter 2011). Hence, more research is needed about the role of family and teacher credentials in the development of children ability.

In this study, we use a rich dataset on family, teacher, and student form China Education Panel Survey (CEPS). CEPS is a large-scale, nationally representative, longitudinal survey starting with the 7th and 9th graders in the 2013-2014 academic year in China. CEPS designs a standard cognitive ability test to measure cognitive ability of each student in the survey. In addition, based on detailed questionnaires about student attitudes, we construct noncognitive ability measures such as confidence, perseverance, and college intention. As we document below, we find that both family inputs and teacher credentials affect student cognitive ability development in China. In addition, we find family plays a critical role in developing student noncognitive ability compared to teacher credentials.

This study makes three contributions to the current literature. First, based on our best knowledge, this is among the first group of studies that evaluate the impact of family inputs and teacher credentials on the development of ability among Chinese middle school students. Such analysis have potential benefits to a country that has the one of the largest education system in the world, China. Specifically, in 2014, China has approximately 3.4 million middle teachers taught nearly 44 million middle school students (China Statistically Yearbook 2015). In addition, the inclusion of both family inputs and teacher credentials reduce the omitted variable problem to a large extent.

Second, the focus on teacher credentials have potential policy implications. In China, teachers are classified into five ranks as an indication of their professional status. The recruitment and promotion are mainly based on education background, years of experience, and current rank. However, it is still an open question whether these credentials, especially the rank, are efficient indicators of teacher quality (Chu et al. 2015). Not only in China, such standards are being used by other countries such as United States. However, a group of scholars and educators believe that the recruitment and promotion of teachers should base on their cognitive ability and effectiveness in the classroom instead of credentials (Walsh 2001). Thus, a better understanding of the relationship between teacher credentials and student ability development can help government assess the effectiveness of the current teacher policies.

Third, our focus on the development of ability is closely connected to the education policy in China. Started from the 21st century, China started to reform the curriculum at the national level with the main purpose of reduce student course load and promote “quality education”. A major document issued by Ministry of Education in 2001 calls for a series of changes such as the de-emphasis of pure “bookish” knowledge and increase of student ability to analyze and solve problems (Ministry of Education 2001).

4.2 Literature Review

By examining the role of family and teacher credentials on individual ability development, our study makes a contribution to the recent literature about the importance of early childhood ability development pioneered by Heckman and co-authors (Cunha et al. 2005, Cunha and Heckman 2007). They emphasize that skill formation is a life cycle process with multiple stages, with both home and school playing a crucial role in developing child’s cognitive and noncognitive abilities. However, family plays a role in this process that is far more important than the role of schools. According to Cunha et al. (2005), there exists critical and sensitive periods of skill formation and remediation. Based on a vast empirical evidence about child ability development, they found that differences in both cognitive and noncognitive ability across different family types appear at early ages and persist. That is, good families with enriched parental environments promote individual ability development while poor families do not.

The importance of family factors such as family size, family income, and parental education on individual achievement has been documented by many other studies. For instance, studies have found that an increased family size has a negative impact on individual’s educational and labor market outcomes (Hanushek 1992, Conley and Glauber 2006). Specifically in China, Rosenzweigh and Zhang (2009) showed that an extra child negatively affect school progress, expected college enrollment, and the health of all children in the family. In addition, Juhn et al. (2015) found that increased family size decrease childhood cognitive abilities and increase behavioral problems (i.e., noncognitive) using data from the National Longitudinal Survey of Youth 1979 (NLSY79) in the United States. In NLSY79, cognitive ability is measured by Peabody Individual Achievement Test (PIATs) in mathematics, reading recognition, and reading comprehension. Noncognitive ability is represented by a Behavioral Problem Index (BPI), which measures particular problems such as antisocial behaviors, anxiety, and hyperactivity.

Family income is another significant predictors of the child's academic achievement. Since Coleman Report highlighted the relationship between family socio-economics status and student achievement in 1966, it has been empirically tested by a series of studies. For instance, using three longitudinal surveys of high school leavers sponsored by the U.S. National Center for Education Statistics (NCES), Acemouglu and Pischke (2001) found that a 10 percent increase in family income is associated with a 1.4 percent increase in the probability of attending a four-year college. Similarly, parental education level is also found to improve individual academic achievement (Holmlund et al. 2011, de Haan 2011). For instance, Carneiro et al. (2011) studied the intergenerational effects of maternal education on children's cognitive and noncognitive ability abilities, grade repetition, and obesity using data from female participates of NLSY79 and their children. The results found maternal education has positive impact on both cognitive ability (i.e., PIAT scores) and behavioral problems (i.e., BPI) of children.

Similar to family background, the causal relationship between teacher quality and student achievement is also established in previous studies (Rockoff 2004, Rivkin et al.

2005, Kane et al. 2006). For instance, Clotfelter et al. (2010) found compelling evidence that teacher credentials such as licensure and certification affect student achievement significantly. Specifically, student achievement is measured by a standard end-of-course tests in five subjects at the high school level in North Carolina. In another study, Chu et al. (2015) evaluated the impact of teacher credentials using a sample of rural school in Ankang Prefecture in Shaanxi province (i.e., one of the poorest regions) in China and found that a teacher with the highest rank (i.e., a credential indicating professional status) has positive impact on student achievement and the impact is heterogeneous base on economic status of students.

Different from studies focusing on student academic test scores or cognitive ability, research about the impact of teach quality on students' noncognitive ability development is very limited (Jackson 2012). In addition to the delivery of content knowledge, teachers also convey value systems and social norms that affect their students' personal characters and social behaviors, which can significantly affect individual future outcome. Based on data on all 9th grade public school students in North Carolina from 2005 to 2011, Jackson (2012) found that teacher quality has causal effect on noncognitive ability measured by absences, suspensions, and on-time grade progression.

In the current study, based on a rich dataset including detailed information about students, parents, teachers, and schools, we add to the previous literature by assessing and comparing the impact of family background and teacher credentials on the development of both cognitive and noncognitive ability among middle school students in China. Our study sheds light on the topic about the multi-stage human capital development and the special emphasize of family inputs compared to school, which has been frequently discussed and theoretically modeled but rarely empirically tested (Cunha et al. 2005).

4.3 Data

Our data come from the China Education Panel Survey (CEPS), a large-scale, nationally representative, longitudinal survey starting with the 7th and 9th graders in the 2013-2014 academic year. The survey sample contains approximately 20,000 students in 438 classrooms of 112 schools in 28 county-level units in mainland China. By conducting different surveys to the sample students, parents, homeroom teachers, main subject teachers, and school officials, CEPS aims to explain the linkages between individuals' educational outcomes and multiple inputs including family, school, community, and social structure.

Commonly used measures of cognitive ability in literature include IQ tests, the Armed Forces Qualification Test (AFQT), and reading, writing, mathematics, and science tests administered by educational institutions. CEPS measures each student's cognitive ability by administrating a national standard cognitive ability test. The test is designed by a special team in CEPS following several important principles. Specifically, it aims to test students' thinking logic and problem-solving ability instead of "bookish" knowledge that needs to be remembered. In addition, the test questions are irrelevant to the specific subject knowledge and background scenarios are familiar to students at this age. The questions falls into three dimensions, including language, graph and space, and calculation and logic. After a pre-test in two middles schools in Shanghai and Zhengzhou and three rounds of revision, the final cognitive ability test contains 20 questions for grade 7 and 22 questions for grade 9. Each test has a 15 minutes time limit and is administered during normal class time. Final cognitive ability score for each student is calculated based on the three-parameter logistic (i.e., 3PL) model in Item Response Theory (Baker and Kim 2004).

Similar to cognitive ability, there are many different ways to measure noncognitive ability. Some commonly used indices of noncognitive ability include Rotter's measure of locus of control (Rotter 1966), the Rosenberg self-esteem scale (Rosenberg 1965), the Five-Factor Model of Personality (Muller and Plug 2006), and emotional intelligence (Goleman 2000).

In our study, we construct four noncognitive ability measures including confidence, college intention, perseverance, and school behavior. Confidence is measured with the question "Do you have confidence about your future?" Students select their response from "not at all", "a little", "to a high extent", and "to a very high extent". Individuals will be categorized in the high-confidence group if they choose "to a very high extent" (i.e., *Confidence=1*). Otherwise, *Confidence=0.* Similarly, college intention is measure by the response to "What is your expectation about your future education?" We set *College=1* if an individual expect to have an education degree above (including) college. Otherwise, *College=0*. Perseverance is measured by several statements including "I will still go to school even I feel uncomfortable", "I will try my best to do homework even if I am not interested in that subject", and "I will continuously try to finish my homework even if it takes a long time". Students select their response from "strongly disagree", "disagree", "agree", and "strongly agree". We set *Perseverance=1* if the responses to all these questions are "strongly agree". Otherwise, *Perseverance=0.* School behavior is measured with two statement: "I am often late for school" and "I often skip classes". We define that individuals have problematic school behavior (i.e., *Behavior=1*) if they answer "agree" or "strongly agree" to either statements; otherwise, *Behavior=0*.

Table 13 presents descriptive statistics of student ability. Table 13 shows that the average cognitive ability test score is close to standard normal distribution after 3PL model, with the mean being zero and the standard deviation being 0.86. In addition, students with confidence, college intention, and perseverance account for 37%, 64%, and 26% of the sample, respectively. Students with behavior problems such as late for school or skip class account for 6%. Moreover, Table 13 shows that 51% of the participated students are male.

Our family background variables can be fit into five categories, including parental relationship, family investment, family size, parental education, and family income. Specifically, parental relationship is represented by whether parents fight frequently (yes=1). Family investment variables include whether there are a lot of books at home (yes=1), whether the students has his/her own desk (yes=1), and computer with internet connection (yes=1). Parental education variables are measured by whether father or mother has a high school degree. Family size is indicated by number of siblings. Since the survey does not have direct information about family annual income, we use whether a family receives government subsidy or has a flush toilet as indicators of family wealth.

Table 13 shows that approximately 10% of student parents fight frequently. In addition, 39% of student families have many books, and 61% have computer with internet connection, and 80% of students have their own desk at home. Average number of siblings is less than one. Moreover, approximately 28% of student mothers and 35% of fathers have a high school degree. Students from families that receive government subsidy and have flush toilet account for 11% and 75% of the sample, respectively, and the correlation between them is -0.31.

Table 13. Descriptive statistics

Variable	Observation	Mean	S.D.	Min	Max
Student					
Cognitive ability	19062	0.00	0.86	-2.03	2.71
Confidence (Yes=1)	18903	0.37	0.48	0	1
College (Yes=1)	18310	0.26	0.44	0	1
Perseverance (Yes=1)	18903	0.64	0.48	0	1
Behavior problem (Yes=1)	18912	0.06	0.24	0	1
Male (Yes=1)	18793	0.51	0.50	0	1
Family background					
Relationship (Fight frequently=1)	18423	0.10	0.30	0	1
Many books (Yes=1)	19007	0.39	0.49	0	1
Own desk (Yes=1)	18679	0.80	0.40	0	1
Computer and internet (Yes=1)	18831	0.61	0.49	0	1
Number of siblings	18776	0.74	0.83	0	5
Mother education (High school =1)	18710	0.28	0.45	0	1
Father education (High school=1)	18675	0.35	0.48	0	1
Poor (Yes=1)	17824	0.11	0.31	0	1
Flush toilet (Yes=1)	18022	0.75	0.43	0	1
Teacher human capital					
Rank (High=1)	18968	0.64	0.48	0	1
Experience	18898	15.63	7.55	0	38
College (Yes=1)	19016	0.45	0.50	0	1
Male (Yes=1)	19062	0.37	0.48	0	1
School characteristics					
Equipment (Yes=1)	18893	0.85	0.36	0	1
Class size	19062	51.62	8.55	25	70
Student teacher ratio	18437	12	4	3	27
Location (Urban=1)	19062	0.39	0.49	0	1

Note: 1. Teachers are classified to high rank if they are level one or higher rank teachers.
2. In school characteristics variables, equipment indicates whether all the classrooms are equipped with multi-media teaching equipment.

In addition to family variables, we include in our analysis several indicators of teacher credentials such as teacher rank, years of experience, and educational background. In china, teacher are classified into five ranks as an indication of their professional status (in ascending order of prestige), from “level three”, “level two”, “level one”, “senior”, to “senior primary”. We define teachers with *rank=1* if they are level one or higher rank teachers. Otherwise, *rank=0*. Table 13 shows that approximately 64% of the teachers has a rank above “level one” and the average years of experience is 15. Approximately 45% of the teachers have a college degree, and 37% are male.

Table 13 also displays school characteristics such as equipment, class size, student teacher ratio, and school location. Specifically, equipment indicates whether all the classrooms are equipped with multi-media teaching equipment. Statistics shows that 85% of the schools are equipped with multi-media equipment, the average class size is 51, and the average student teacher ratio is 12. On average, 39% of the school are in urban areas.

4.4 Basic estimation result

Our basic estimation is based on the following equation,

$$y_{ijk} = \alpha + Family_{ijk}\theta + Teacher_j\gamma + School_k\rho + u_{ijk},$$

where y_{ijk} refers to the ability of student i with teacher j at school k, $Family_{ijk}$ contains individual i′s family back ground variables such as number of sibling and parental education, $Teacher_j$ is a vector of variables that describe teacher j′s credentials such as rank and years of experience, $School_k$ controls for school level variables such as equipment and class size, and u_{ijk} is the error term.

Cognitive ability effects of family and teacher credentials by grade level

The basic results for the impact of family and teacher credentials on student cognitive ability is reported in Table 14. In the first three models, error terms are clustered at the class level to control for any correlation of errors associated with the common experience of students in a specific class. In the last model, we apply the feasible general least square estimation (FGLS), which is more efficient than clustered estimation. Specifically, Model 1 and 2 control for family characteristics and teacher human capital variables, separately, and Model 3 and 4 control for both. Since Model 3 and 4 are more general, and estimation results are highly consistent with the previous two models, we will focus our discussion on them. Begin with the results from family background variables, we find that if a student's parents fight frequently, the students will have a lower cognitive ability test score. In addition, a large number of books and computers with internet connection in a student's house also significantly affect the individual's cognitive ability. Moreover, own a desk at home increases student cognitive ability test scores by 0.069 to 0.090.

Table 14 also shows that number of siblings has a significant negative impact on cognitive ability. Specifically, if the number of siblings increases by one, cognitive ability test scores will decrease by 0.082 to 0.096. The result is consistent with previous literatures about the quantity and quality trade off with respect to family size. Increase quantity will reduce the available resource on each child and potentially affect their future outcomes (Juhn et al. 2015).

We also observe that both mother and father's education background positively impact student cognitive ability, and father's impact is stronger. Students whose father has a high school diploma or a higher degree have cognitive ability 0.079 higher than those who do not. Holmlund et al. (2011) reviewed studies that aim to estimate the causal link between parental education and their children. For studies that found significant impacts from parents, the influence of mother's education is somewhat larger than fathers. However, Behrmand and Rosenzweig (2002) found that father's schooling is the most important.

Table 14 Cognitive ability effects of family and teacher credentials

	Cluster			FGLS
	Family	Teacher	Both2	Both2
Family				
Relationship (Fight frequently=1)	-0.062***		-0.073***	-0.050**
Many books (Yes=1)	0.207***		0.185***	0.176***
Own desk (Yes=1)	0.139***		0.090***	0.069***
Computer and internet (Yes=1)	0.145***		0.095***	0.073***
Number of siblings	-0.120***		-0.096***	-0.082***
Mother education (High school =1)	0.079***		0.045**	0.050***
Father education (High school=1)	0.104***		0.079***	0.079***
Poor (Yes=1)	-0.175***		-0.146***	-0.128***
Flush toilet (Yes=1)	0.071***		0.009	0.025
Teacher human capital				
Rank		0.209***	0.148***	0.146***
Experience		0.007*	0.004	0.004***
College (Yes=1)		0.182***	0.111**	0.118***
Male (Yes=1)		-0.099**	-0.058	-0.061***
School characteristics		Yes	Yes	Yes
Individual characteristics	Yes	Yes	Yes	Yes
Constant	-0.234***	-0.246*	-0.260*	-0.245***
N	15773	17854	14982	14982
Adjusted R-sq	0.122	0.098	0.143	
F	51.148	23.775	30.439	

Note: 1. *** indicates significance at 1% level, ** at 5% level, and * at 10% level.
2. Error are clustered by classroom in the cluster estimation. The last column presents the feasible GLS estimation.
3. Personal characteristics variables include gender.
4. School characteristics variables include equipment (i.e., whether all the classrooms are equipped with multi-media teaching equipment), class size, student teacher ratio, and school location (i.e., urban=1).

Emerging from Table 14 is the significant impact of family income. For instance, a students from a family receives government subsidy has a cognitive ability test score

0.128 to 0.146 lower compared to a student who does not. Have a flush toilet does not have a significant impact on cognitive ability. The significance of family income is consistent with previous studies (e.g., Acemoglu and Pischke 2001).

In addition to family background, teacher human capital are also predictive of student cognitive ability. Table 14 shows that teacher rank has a positive and significant impact on the achievement of student cognitive ability. Specifically, a higher rank (i.e., level one or above) increases student ability test scores by 0.146 to 0.148. The result is highly consistent with another study about China by Chu et al. (2015), who found that teacher rank increases student achievement by 0.23 standard deviations.

We find one more year of experience has significant impact on cognitive ability based on the result of FGLS, but the magnitude is small. Based on previous literature, the impact of teacher experience on student achievement, commonly measured by subject test scores, is nonlinear, with the highest impact in the early years (Clotfelter et al. 2006, 2010). However, we do not find any nonlinear impact by disaggregate our experience variable. We conjecture the possible reason is that the nonlinear impact of experience on ability is less obvious compared to test score.

Table 14 shows that having a college or higher degree is predictive of higher cognitive ability compared to having a teacher without a college degree. The results suggest that a teacher's education background is important in developing students' cognitive ability. In addition, an interesting result is the significant and negative impact of male teachers. Based on previous literature, the negative impact appears because of the negative interactions between male teachers and female students (Clotfelter et al. 2010).

We divide our sample by student population and find that the negative impact from male teacher disappears.

To summarize, our basic estimation results indicate family background such as family environment, family size, parental education, and family income are all predictive of student cognitive ability. With respect teacher credentials, the significant impact occurs with rank, experience, and teacher's education background. It seems teacher's rank if the most predictive among all human capital measures.

Noncognitive ability effects of family and teacher credentials

Table 15 presents family and teacher human capital's impact on student noncognitive abilities, including confidence, college intention, perseverance, and behavior problem. For Model 1, Table 15 shows if parents fight frequently, the student is less likely to be confident. Having many books and own desk at home also help the development of confidence. In addition, if number of siblings increase by one, the probability of being confidence will decrease by 0.016. Different from cognitive ability, whether father play a more important role, mother's education level is more important in predicting student's confidence level. Coming from a poor family does not significantly affect students' confidence level. However, having a flush toilet has a significant and negative impact. With respect to teacher credentials, different from our previous estimation that a higher teacher rank promotes cognitive ability, a higher teacher rank actually decreases student confidence level, with the magnitude of 0.042. In addition, compared to female teacher, male teachers decrease students' probability of being confidence by 0.053.

Table 15 Noncognitive ability effects of family and teacher credentials

	Confidence	College	Perseverance	Behavior
Family				
Relationship (Fight frequently=1)	-0.077***	-0.043***	-0.068***	0.037***
Many books (Yes=1)	0.154***	0.134***	0.096***	0.002
Own desk (Yes=1)	0.073***	0.053***	0.030***	-0.013*
Computer and internet (Yes=1)	-0.004	0.004	-0.019**	-0.002
Number of siblings	-0.016***	-0.033***	-0.012**	0.009***
Mother education (High school =1)	0.063***	0.055***	0.01	0.003
Father education (High school=1)	0.016*	0.084***	-0.019*	0.003
Poor (Yes=1)	0.002	-0.041***	-0.003	0.014**
Flush toilet (Yes=1)	-0.038***	-0.020*	-0.038***	0.009*
Teacher human capital				
Rank	-0.042**	0.002	-0.032*	-0.001
Experience	0.001	0.002*	0	0
College (Yes=1)	0.006	0.02	0.001	-0.003
Male (Yes=1)	-0.053***	-0.008	-0.037**	0.006
Personal and school characteristics	Yes	Yes	Yes	Yes
Constant	0.199***	0.467***	0.350***	0.024
N	14896	14896	14539	14922
Adjusted R-sq	0.051	0.102	0.023	0.006
F	32.581	66.494	18.607	4.979

Note: 1. *** indicates significance at 1% level, ** at 5% level, and * at 10% level.
2. All the results are based on cluster estimation, and error terms are clustered by classroom.
3. Personal characteristics variables include gender.
4. School characteristics variables include equipment (i.e., whether all the classrooms are equipped with multi-media teaching equipment), class size, student teacher ratio, and school location (i.e., urban=1).

Model 2 presents the result for college intention. The impact (i.e., signs) from family variables such as family environment, size, parental education, and family income are quite similar to that of cognitive ability. Teacher experience is the only significant variable in predicting student college intention. With regards to perseverance in Model 3, the impact of family variables are quite consistent with Model 1, except that father's education becomes negatively significant. In addition, if a family has a computer with internet connection, the probability of having strong perseverance will decrease by 0.019. The possible reason is that computer is a temptation that prevents students from finishing their work. Again, we do not observe much impact from teacher credentials, with the only significant variable being teacher's rank. Moreover, having a male teacher decreases student's perseverance level.

Model 4 presents the result for behavior problems. Note that R square is very low, less than 1%, indicating that a significant amount of behavior problem cannot be explained by our current explanatory variables. We find that if parent fight frequently, students are more likely to have behavior problems. In addition, more siblings and being in a poor family will increase the probability of having behavior problems.

In summary, compared to cognitive ability, the explanatory power of our variables is limited for noncognitive ability. Overall, we find the role family on the development of noncognitive ability is stronger than teacher credentials among Chinese middle students. Specifically, we observe that parental relationship is important in the development of noncognitive ability. Having parents who fight frequently prevents the formation of desirable qualities. In addition, there is a large tradeoff between quality and quantity with respect to family size. Increased number of siblings has negative impact on preferable noncognitive abilities. Moreover, a poor family negatively affect student's college intention and increase behavior problems. With respect to teacher credentials, different from cognitive ability, we find teacher rank negatively affect student's confidence and perseverance. Male teachers are at a disadvantaged position in developing student noncognitive ability compared to female teachers.

4.5 Further investigation

In this section, we test the robustness of our results with regards to across-school sorting and within-school sorting (Clotfelter et al. 2006). Across-school sorting means students and teachers are not assigned to school randomly. Teachers choose school to teach based on salary, benefits, school location, and student characteristics, and high quality teachers are more likely to end up in school with more advantaged students. Meanwhile, parents who want their children receive high quality education may also choose to move to the district with good schools. Similarly to across-school sorting, within-school sorting happens if teachers and students are not assigned to classrooms randomly. School officials may assign high quality teachers to more advanced students. In addition, parents may also make the most preferable teacher to teach their child by trying to change the class assignment decisions, so call “teacher shopping”.

To deal with the random sorting, we applied three strategies. First, we control for school fixed effects. The inclusion of school fixed effect implicitly control for unobservable characteristics that vary by school. Thus, the coefficients of teacher credentials are only identified by the variations within a school (Kane et al. 2008). However, doing so will wash out all the differences in teacher credential across school. As a result, we consider the result here as a lower bound whereas the result before as an upper bound. Second, we restrict our sample to the classes that are not assigned based on student overall performance or performance for a single subject.[59] Thus, any within school sorting will reduced, if not eliminated. Third, we applied the Hausman Taylor

[59] In the questionnaire, homeroom teachers were asked if all the classes in grade are classified based on students overall performance or a particular subject. We only keep the classes that are classified not based on performance.

estimation, which eliminates school and classroom unobserved effects and recover teacher's coefficients at the same time. We specify parental education, family income, and teacher human capitals as endogenous variables because they are likely to be correlated with unobserved school and class fixed effects in the error term.

Table 16 presents the results of further investigation with school fixed effects. Model 1 presents the results for the whole sample, and Model 2 and 3 present results for grade 7 and grade 9, respectively. With the inclusion of fixed effects, the impact of family background variables remain highly consistent with previous estimation. However, all teacher human capital variables become insignificant in Model 1. When we look at the analysis result by grade, teacher's rank and experience become significant only for grade 7 students, indicating the importance of early intervention. Overall, with the inclusion of school fixed effects, both the impact from family background and teacher human capital become smaller because variations between schools are removed.

Table 16 Cognitive ability effects of family and teacher credentials with school fixed effects

	Whole sample	Grade 7	Grade 9
Family			
Relationship (Fight frequently=1)	-0.056***	-0.047	-0.056**
Many books (Yes=1)	0.144***	0.118***	0.157***
Own desk (Yes=1)	0.041**	0.031	0.056**
Computer and internet (Yes=1)	0.024	0.035	-0.002
Number of siblings	-0.032***	-0.021	-0.036***
Mother education (High school =1)	0.025	0.007	0.040*
Father education (High school=1)	0.077***	0.048**	0.078***
Poor (Yes=1)	-0.113***	-0.084***	-0.127***
Flush toilet (Yes=1)	0.034*	0.035	0.057**
Teacher human capital			
Rank	0.063	0.094*	0.033
Experience	0.003	0.005*	-0.003
College (Yes=1)	0.064	-0.045	-0.05
Male (Yes=1)	-0.075**	-0.056	-0.156***
Personal and school characteristics	Yes	Yes	Yes
School dummies	Yes	Yes	Yes
Constant	-0.42	0.403**	1.344**
N	14982	7751	7231
Adjusted R-sq	0.235	0.249	0.289

Note: 1. *** indicates significance at 1% level, ** at 5% level, and * at 10% level.
2. All the results are based on cluster estimation, and error terms are clustered by classroom.
3. Personal characteristics variables include gender.
4. School characteristics variables include equipment (i.e., whether all the classrooms are equipped with multi-media teaching equipment), class size, student teacher ratio, and school location (i.e., urban=1).

Table 17 presents the results of two strategies that solve both across-school and within-schooling sorting (i.e., unobserved school and unobserved classroom fixed effects). Model 1 displays the result with school fixed effects and subsample restriction, and model 2 shows the result of Hausman Taylor estimation. For family background variable, the results for both strategies are highly consistent with Table 14. For teacher human capital, Model 1 does not find significant impact from teacher's rank for the whole sample. However, when we split the sample into grades, teacher's rank has a significant impact on cognitive ability for Grade 7 students, which is consistent with result in Table 16.[60] Model 2 shows that both teacher's rank and education background have significant impact on cognitive ability. Overall, results of further investigation in which endogeneity issues are controlled for are in line with our basic estimation.

Table 17 Cognitive ability effects of family and teacher credentials – Endogeneity

	School fixed effect and subsample restriction	Hausman Taylor Estimation
Family		
Relationship (Fight frequently=1)	-0.044*	-0.040**
Many books (Yes=1)	0.157***	0.126***
Own desk (Yes=1)	0.037	0.037**
Computer and internet (Yes=1)	0.016	0.009
Number of siblings	-0.029**	-0.024***
Mother education (High school =1)	0.023	0.023
Father education (High school=1)	0.082***	0.053***
Poor (Yes=1)	-0.101***	-0.092***
Flush toilet (Yes=1)	0.038*	0.033*
Teacher human capital		
Rank	-0.007	0.914**
Experience	0.004	0.008
College (Yes=1)	-0.03	0.883***
Male (Yes=1)	-0.095**	-0.011
Personal and school characteristics	Yes	Yes
School dummies	Yes	Yes
Constant	0.093	-1.577***
N	10813	14982
Adjusted R-sq	0.255	

Note: 1. *** indicates significance at 1% level, ** at 5% level, and * at 10% level.
2. Personal characteristics variables include gender.
3. School characteristics variables include equipment (i.e., whether all the classrooms are equipped with multi-media teaching equipment), class size, student teacher ratio, and school location (i.e., urban=1).

[60] The estimation result for different grade is not presented here due to space limit.

4.6 Conclusion

In this study, we investigate the impact of family and teacher credentials on the development of cognitive and noncognitive ability among Chinese middle school students. Our analysis is based on data form China Education Panel Survey (CEPS), which provides detailed information about students, parents, teachers, and schools. Our cognitive ability measure is obtained from a national standard test designed by CEPS, and our noncognitive ability indicators include confidence, college intention, perseverance, and behavior problems.

Our results show that both family and teacher credentials have significant impact on the development of cognitive ability. For family variables, we find a good parental relationship is important for the development of cognitive ability. Better environment such as the presence of books and study desks at home, higher parental education level, and family income will also improve student cognitive ability. In addition, increased number of siblings will decrease student cognitive ability. With respect to teacher credentials, we find teacher rank is the most significant variable in predicting student cognitive ability, even after we control for between- and within-school sorting.

With regards to noncognitive ability, we find family background plays a more critical role compared to teacher credentials. Parental relationship is critical in developing desirable noncognitive abilities. A better environment also promotes noncognitive ability. Increased number of siblings decrease the probability of being confidence, having a college intention, or having strong perseverance, but increase the probability of having behavior problems. For teacher credentials, we find an unexpected and negative impact from teacher rank. Thus, we conjecture that the development of student noncognitive ability may be influenced by teacher’s soft skills instead of quality. We will leave this question for future studies.

CHAPTER 5

CONCLUSION AND OUTLOOK

This dissertation evaluates the impact and the development of human capital. We use cognitive and noncognitive ability as our main indicators of human capital. Chapter 2 evaluates the relationship between human capital and leadership. The major conclusion of Chapter 2 is that both cognitive and noncognitive abilities are important predictors of leadership, with the most important ability measures being problem-solving and perseverance. Specifically, a one unit increase in problem-solving test scores will increase the probability of being a leader by 0.273 to 0.3. Individuals with high levels of perseverance have an 8.4 to 9.0 percentage point higher chance of being in a leadership position compared to those with low levels of perseverance.

By evaluating factors that affect MOOC demand, Chapter 3 finds that an increased unemployment rate and internet speed increase MOOC demand in both OECD countries and in China. In OECD countries, we also observe significant impact form the percent of college and high school graduates. Different from OECD countries, we find wage has positive impact on MOOC demand in China. We do not observe any significant impact from tuition per capita, which indicates that there is no strong complement or substitute effects between MOOC and traditional education.

Chapter 4 finds that both family background and teacher credentials have significant impact on student ability development. Notably, we find teacher rank is the most effective predictor of teacher quality, indicating the current teacher evaluation system in China can differentiate high quality teachers, at least to some extent. This is fully consistent with previous study in China conducted by Chu et al. (2015). With respect to noncognitive ability, family plays a positive and a more crucial role compared to teacher human capital. Especially, we find parental relationship is important in shaping student noncognitive ability.

Our research can be enriched from the following aspects. First, there is no consensus in literature about whether the current existing measures of ability can capture one's true inner ability, whether they are correlated with the uncaptured part of the ability, and to what extent one's inner ability can be changed because of the environment (e.g., workplace). Research in this area will significantly impact the empirical model specification and estimation methodology for any economic studies. Second, our MOOC study is restricted by the availability of data. Since MOOC took off in 2012, some local platforms based on Chinese language appeared in 2013 or 2014. It would be interesting to compare the difference between the demand for international platforms and local Chinse platforms in China. As data become available, a systematic local platform analysis can be conducted. Third, at the current stage, CEPS only evaluates student cognitive ability based on a national standard tests. However, to the extent that ability is different from subject test scores in midterm or final exams, we can furfure analyze the impact of family and teacher credential on student test scores such as Chinese, math, or science.

APPENDIX A

FIRST STAGE IV

Table 7. First stage IV regression

	Numeracy	Literacy	Problem solving	Education
Male	0.173***	0.022	0.052*	-0.008
Experience	-0.007***	-0.006***	-0.010***	-0.011***
Married	0.164***	0.109***	0.095***	0.096**
Either parent had a college degree	0.192***	0.183***	0.116***	0.189***
Number of kids	-0.038***	-0.031**	-0.021	-0.023
Perseverance	0.065*	0.034	0.044	0.060
Openness to learning	-0.028	-0.013	-0.011	0.024
Social trust	0.124***	0.100***	0.126***	0.170***
Parents foreign born	0.066	0.097*	-0.026	0.045
Test language same as language usually spoken at home	0.280***	0.239***	0.213***	0.195**
Foreign years	-0.004	-0.002	-0.007	0.022**
Foreign years squared	0	0	0	0
Looking after children	0.001	0	-0.043	-0.114**
Interrupted by other activities	-0.019	0	0.009	-0.008
Parents foreign born*Foreign years	-0.004	-0.020***	-0.009*	-0.012*
Constant	2.421***	2.682***	2.761***	0.405***
Number of observations	751	751	751	751
R^2	0.251	0.252	0.247	0.252
Adjusted R^2	0.234	0.235	0.230	0.234
F-statistics	16.861	14.711	13.41	25.308

Note: *** denotes significance at 1%, ** denotes significance at 5%, and * denotes significance at 10%.

APPENDIX B

CHINA LOCAL PLATFORMS

Table 12. Factors affect MOOC demand in local platforms in China

	FE	Dummy	Trend
Tuition per capita	0.274	-0.439	-1.951
Unemployment	1.491*	-0.066	1.388**
Wage	17.988***	1.851**	1.466
Internet speed	2.387*	0.171	-3.297***
Percent of college graduates	-0.035	0.003	-0.057
Percent of high school graduates	-0.078	-0.024	-0.017
Internet users	-5.225	-1.449	-6.76
Year dummy 2013		2.722***	
Year dummy 2014		3.331***	
Trend			3.526***
Constants	-21.374	3.573	24.849
N	91	91	91
R-squared	0.755	0.993	0.881
F	44.828	1342.026	87.873

Note: 1. Number of observation for local platforms is smaller than international platforms because local platforms appear one year after international platforms.
2. In our three-year local platform model, controlling for year dummies is equivalent to controlling for trend and trend^2. Thus, nonlinear trend specification is not presented here.
3. *** represents significance at 1% significant level, ** represents significance at 5% significant level, and * represents significance at 10% significance level.

REFERENCES

Acemoglu, D. and J.S. Pischke. 2001. Changes in the Wage Structure, Family Income, and Children's Education. *European Economic Review* 45: 890-904.

Akamai. 2015. State of the Internet Report. https://content.akamai.com/PG5641-Q4-2015-SOTI-Connectivity-Report.html

Arvey, R.D., Z. Zhang, B.J. Avolio, and R.F. Krueger. 2007. Development and Genetic Determinants of Leadership Role Occupancy among Women. *Journal of Applied Psychology* 92(3): 693-706.

Ayyar, R. 2016. Flipkart Hires without Interview. http://timesofindia.indiatimes.com/tech/tech-news/Flipkart-hires-without-interviews/articleshow/50756104.cms

Baker, F.B. and S. Kim. 2004. Item Response Theory: Parameter Estimation Techniques. 2nd ed. New York: Dekker.

Barber, M., K. Donnelly, and S. Rizvi. 2013. An Avalanche is Coming: Higher Education and the Revolution Ahead. Institute for Public Policy Research. London.

Bass, S.A. 2014. Simple Solutions to Complex Problems – MOOCs as a Panacea? *The Journal of General Education* 63(4): 256-268.

Behrman, J. R., and M. R. Rosenzweig. 2002. Does Increasing Women's Schooling Raise the Schooling of the Next Generation? *American Economic Review* 92 (March): 323–34.

Borghans, L., B. Well, and B.A. Weinberg. 2008. Interpersonal Styles and Labor Market Outcomes. *The Journal of Human Resources* 43: 815-858.

Breslow, L., D.E. Pritchard, J. DeBoer, G.S. Stump, A.D. Ho, and D. Seaton. 2013. Studying learning in the worldwide classroom: Research into edX's First MOOC. *Research & Practice in Assessment* 8: 13–25.

Brunello, G. and M. Schlotter. 2011. Non Cognitive Skills and Personality Traits: Labour Market Relevance and their Development in Education and Training Systems. IZA DP No. 5747.

Carneiro, P., C. Meghir, and M. Parey. 2013. Maternal Education, Home Environments, and the Development of Children and Adolescents. *Journal of the European Economic Association* 11: 123-160.

Card, D., 1999. The causal effect of education on earnings. In: Ashenfelter, O., Card, D. (Eds.), Handbook of Labour Economics, vol. 3. North-Holland, Amsterdam, 1801 – 1863.

Carson, J.B., P.E. Tesluk, and J.A. Marrone. 2007. Shared Leadership in Teams: An Investigation of Antecedent Conditions and Performance. *Academy of Management Journal* 50: 1217–1234.

Cawley, J., J.J. Heckman, and E. Vytlacil. 2001. Three Observations on Wages and Measured Cognitive Ability. *Labour Economics* 8(4): 419–442.

Christensen, G., A. Steinmetz, B. Alcorn, A. Bennet, D. Woods, and E.J. Emmanuel. 2013. The MOOC Phenomenom: Who Takes Massive Open Online Courses and Why? University of Pennsylvania. Working Paper.

Choi, H., and H. Varian. 2012. Predicting present with google trends. *Economic Record* 88(S1): 2-9.

Chui, M. 2015 Coursera Takes Aim at Unemployment. http://www.mckinsey.com/industries/social-sector/our-insights/coursera-takes-aim-at-unemployment

Chu, J.H., P. Loyalka, P. Loyalka, J. Chu, Q. Qu, Y. Shi, and G. Li. 2015. The Impact of Teacher Credentials on Student Achievement in China. *China Economic Review* 36: 14-24.

Clotfelter, C.T., H.F. Ladd, and J.L. Vigdor. 2006. Teacher-Student Matching and the Assessment of Teacher Effectiveness. *Journal of Human Resources* 41: 778-820.

Clotfelter C.T., H.F. Ladd, and J.L. Vigdor. 2010. Teacher Credentials and Student Achievement in High School. A Cross-Subject Analysis with Student Fixed Effects. *Journal of Human Resources* 45: 655-681.

Conley, D., and R. Glauber. 2006. Parental Educational Investment and Children's Academic Risk: Estimates of the Impact of Sibship Size and Birth Order from Exogenous Variation in Fertility. *Journal of Human Resources* 41: 722-737.

Cobb-Clark, D., and S. Schurer. 2012. The Stability of Big-Five Personality Traits. *Economics Letters* 115: 11-15.

Coleman, J. 1990. Foundations of Social Theory. 1st Edition. Belknap Press.

Connelly, M.S., J.A. Gilbert, S.J. Zaccaro, K.V. Threlfall, M.A. Marks, and M.D. Mumford. 2000. Exploring the Relationship of Leadership Skills and Knowledge to Leader Performance. *The Leadership Quarterly* 11(1): 65-86.

Cunha, F., J. Heckman, L. Lochner, and D.V. Masterov. 2005. Interpreting the Evidence on Life Cycle Skill Formation. NBER Working Paper No. 11331.

Cunha, F. and J. Heckman. 2007. The Technology of Skill Formation. *The American Economic Review* 97(2): 31-47.

Daly, M., M. Egan, and F. O'Reilly. 2015. Childhood General Cognitive Ability Predicts Leadership Role Occupancy across Life: Evidence from 17,000 Cohort Study Participants. *The Leadership Quarterly* 26: 323-341.

Deary, I.J., L.J. Whalley, H. Lemmon, J.R. Crawford, and J.M. Starr. 2000. The Stability of Individual Differences in Mental Ability from Childhood to Old Age: Follow-up of the 1932 Scottish Mental Survey. *Intelligence* 28: 49–55.

de Haan, M. 2011. The Effect of Parents' Schooling on Child's Schooling: A Nonparametric Bounds Analysis. Journal of Labor Economics 29: 859-892.

Dinh, J.E., R.G. Lord, W.L. Gardner, J.D. Meuser, R.C. Liden, and J. H. 2014. Leadership Theory and Research in the New Millennium: Current Theoretical Trends and Changing Perspectives. *The Leadership Quarterly* 25: 36-62.

Dirks, K.T., and D.L. Ferrin. 2002. Trust in Leadership: Meta-Analytic Findings and Implications for Research and Practice. *Journal of Applied Psychology* 87(4): 611-628.

DiSalvo, B., P.K. Roshan, and B. Morrison. 2016. Information Seeking Practices of Parents: Exploring Skills, Face Threats and Social Networks. CHI Conference on Human Factors in Computing Systems, California, USA.

Dries, N., and R. Pepermans. 2012. How to Identify Leadership Potential: Development and Testing of A Consensus Model. *Human Resource Management* 51(3): 361-385.

Emanuel, E.J. 2013. Online Education: MOOCs taken by Educated Few. *Nature* 503: 342.

Eren, O., and S. Ozbeklik. 2013. The Effect of Noncognitive Ability on the Earnings of Young Men: A Distributional Analysis with Measurement Error Correction. *Labour Economics* 24: 293-304.

Fleisher, B., H. Li, and M. Zhao. 2009. Human Capital, Economic Growth, and Regional Inequality in China. *Journal of Development Economics* 92(2): 215-231.

Gibbons, R., and M. Waldman. 2006. Enriching a Theory of Wage and Promotion Dynamics inside Firms. *Journal of Labor Economics* 24(1): 59-107.

Ginsberg, J., M.H. Mohebbi, R.S. Patel, L. Brammer, M.S. Smolinski, and L. Brilliant. 2009. Detecting Influenza Epidemics Using Search Engine Query Data. *Nature* 457(7232): 1012-1014.

Glewwe, P. and H.G. Jacoby. 2004. Economic Growth and the Demand for Education: Is there A Wealth Effect? *Journal of Development Economics* 74(1): 33-51.

Goleman, D. 2000. Leadership that Gets Results. *Harvard Business Review* March-April: 2-17.

Grainger, B. 2013. Massive Open Online Course (MOOC) Report 2013. University of London International Programmes.

Griffiths, R., M. Chingos, C. Mulhern, and R. Spies. 2014. Interactive Online Learning on Campus. ITHAKA.

Hanushek, E. 1992. The Trade-off between Child Quantity and Quality. *Journal of Political Economy* 100: 84-117.

Heckman, J.J., and Y. Rubinstein. 2001. The Importance of Noncognitive Skills: Lessons from the GED Testing Program. *The American Economic Review* 92(1): 145-149.

Heckman, J.J., J. Stixrud, and S. Urzua. 2006. The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcome and Social Behavior. *Journal of Labor Economics* 24(3): 411-482.

Hersey, P., and K. Blanchard. 1988. Management of Organizational Behavior. Englewood Cliffs, New Jersey: Prentice Hall.

Herrnstein, R.J. and C.A. Murray. 1994. *The Bell Curve: Intelligence and Class Structure in American Life*. New York: Free Press.

Holmlund, H, M. Lindahl, and E. Plug. 2011. The Causal Effect of Parent's Schooling on Children's Schooling: A Comparison of Estimation Methods. *Journal of Economic Literature* 49: 615-651.

Hooghe, M., S. Marien, and T. de Vroome. 2012. The Cognitive Basis of Trust. The Relation between Education, Cognitive Ability, and Generalized and Political Trust. *Intelligence* 40: 604-613.

Jackson, K. 2012. Non-cognitive Ability, Test Scores, and Teacher Quality: Evidence from 9th Grade Teachers in North Caroline. NBER Working Paper No. 18624.

Judge, T.A., and J.E. Bono. 2000. Five-factor Model of Personality and Transformational Leadership. *Journal of Applied Psychology* 85: 751–769.

Judge T.A., R.L. Klinger, and L.S. Simon. 2010. Time Is on My Side: Time, General Mental Ability, Human Capital, and Extrinsic Career Success. *Journal of Applied Psychology* 95(1): 92-107.

Juhn, C. Y. Rubinstein, and C.A. Zuppann. 2015. The Quantity-Quality Trade-off and the Formation of Cognitive and Non-cognitive Skills. NBER Working Paper 21824.

Kane, T. J. Rockoff, and D. Staiger. 2006. What Does Certification Tell Us About Teacher Effectiveness? Evidence from New York City. *Economics of Education Review* 27: 615-631.

Krause, S., and C. Lowe. 2014. Invasion of the MOOCs: The promise and perils of massive open online courses. San Francisco: Parlor Press.

Krueger, A.B. 2008. Unemployment in a Volatile Economy: How to Secure Families and Build Opportunity. Hearing of the Committee on Health, Education, Labor, and Pensions. United States Senate.

Kuhn, P., and C. Weinberger. 2005. Leadership Skills and Wages. *Journal of Labor Economics* 23: 395-436.

Lazear, E.P. 2010. Leadership: A Personnel Economics Approach. NBER Working Paper 15918.

Lequerica A. 2016. MOOCs for Credit https://www.class-central.com/report/moocs-for-credit/#take-a-mooc

Lindqvist, E., and R. Vestman. 2011. The Labor Market Returns to Cognitive and Noncognitive Ability: Evidence from the Swedish Enlistment. *American Economic Journal: Applied Economics* 3(1): 101-128.

McGaughey, R.E., and K.H. Mason. 1998. The Internet as a Marketing Tool. *Journal of Marketing Theory and Practice* 6(3): 1-11.

McGuire, R. 2014. The Best MOOC Provider: A Review of Coursera, Udacity, and Edx. http://www.skilledup.com/articles/the-best-mooc-provider-a-review-of-coursera-udacity-and-edx

McGrew, K. S. 2009. CHC theory and the Human Cognitive Abilities Project: Standing on the Shoulders of the Giants of Psychometric Intelligence Research. *Intelligence* 37: 1–10.

Miller, C.C. 2014. Why the U.S. Has Fallen Behind in Internet Speed and Affordability. http://www.nytimes.com/2014/10/31/upshot/why-the-us-has-fallen-behind-in-internet-speed-and-affordability.html?_r=0

Mincer, J. 1974. Schooling, Experience and Earnings. New York: National Bureau of Economic Research.

Ministry of Education of the PRC. 2001. Guidelines for Curriculum Reform in Basic Education. Ministry of Education, Beijing.

Moon, B.J. 2004. Consumer Adoption of the Internet as an Information Search and Product Purchase Channel: Some Research Hypotheses. *International Journal of Internet Marketing and Advertising* 1(1): 104-118.

Muller, G. and E, Plug. 2006. Estimating the Effect of Personality on Male and Female Earnings. *Industrial and Labor Relations Review* 60 (1): 3-22.

Mumford, M.D., S.J. Zaccaro, F.D. Harding, T.O. Jacobs, and E.A. Fleishman. 2000. Leadership Skills for a Changing World: Solving Complex Social Problems. *The Leadership Quarterly* 11(1): 11-35.

Mumford, T.V., M.A. Campion, and F.P. Morgeson. 2007. The Leadership Skill Strataplex: Leadership Skill Requirement across Organizational Levels. *The Leadership Quarterly* 18: 154-166.

Murnane, R.J., J.B. Willett, and F. Levy. 1995. The Growing Importance of Cognitive Skills in Wage Determination. *The Review of Economics and Statistics* 77(2): 251-266.

Navarro, P. 2015. How Economics Faculty Can Survive (and Perhaps Thrive) in a Brave New Online World. *Journal of Economic Perspectives* 29(4): 155-176.

OECD. 2011. The Well-being of Nations: The Role of Human and Social Capital. OECD Publishing.

OECD. 2013. OECD Skills Outlook 2013: First Result from the Survey of Adult Skills. OECD Publishing.

OECD. 2014. PISA 2012 Results: Creative Problem-Solving. OECD Publishing.

OECD. 2014. Access Network Speed Tests. OECD Digital Economy Papers, No. 237, OECD Publishing.

OECD. 2015. OECD Digital Economy Outlook 2015, OECD Publishing.

Perna, L., A. Ruby, R. Boruch, N. Wang, J. Scull, C. Evans, and S. Ahmad. 2013. The life Cycle of A Million MOOC Users. http://www.gse.upenn.edu/pdf/ahead/perna_ruby_boruch_moocs_dec2013.pdf

Peterson, R.A., and M.C. Merino. 2003. Consumer Information Search Behavior and the Internet. *Psychology & Marketing* 20(2):99-121.

Ree, H. M., and T.R. Carretta. 2002. g2K. *Human Performance* 15: 3–24.

Rivkin, S.G., E.A. Hanushek, and J.F. Kain. 2005. Teachers, Schools and Academic Achievement. *Econometrica* 79: 418-458.

Rosen, S. 1982. Authority, Control, and the Distribution of Earning. *Bell Journal of Economics* 13(2): 311-323.

Rosenberg, M. 1965. Society and the Adolescent Self-image. Princeton. NJ: Princeton University Press.

Rotter, J.B. 1966. Generalized Expectancies for Internal Versus External Control of Reinforcement. *Psychological Monographs: General and Applied* 80(1): 1-28.

Rockoff, J.E. 2004. The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data. *American Economic Review Paper and Proceedings* 94: 247-252.

Rosenzweig, M.R., and J. Zhang. 2009. Do Population Control Policies Induce More Human Capital Investment? Twins, Birth Weight and China's One-child Policy. *Review of Economic Studies* 76: 1149-1174.

Schalke, D., M. Brunner, C. Geiser, F. Preckel, U. Keller, M. Spengler, and R. Martin. 2013. Stability and Change in Intelligence from Age 12 to Age 52: Results From the Luxembourg MAGRIP Study. *Developmental Psychology* 49: 1529-1543

Schuwer, R., I. Gil-Jaurena, C.H. Aydin, E. Costello, C. Dalsgaard, M. Brown, D. Jansen, and A. Teixeira. 2015. Opportunities and Threats of the MOOC Movement for Higher Education: The European Perspective. *International Review of Research in Open and Distributed Learning* 16(6): 20-38.

Segal, C. 2013. Misbehavior, Education and Labor Market Outcomes. *Journal of the European Economic Association* 11(4): 743–779.

Shah, D. 2014. Online Courses Raise Their Game: A Revie of MOOC Stats and Trends in 2014. https://www.class-central.com/report/moocs-stats-and-trends-2014/

Shah, D. 2015. Less Experimentation, More Iteration: A Review of MOOC Stats and Trends in 2015. https://www.class-central.com/report/moocs-stats-and-trends-2015/

Shah, D. 2015. How Coursera Cracked The Chinese Market. https://www.class-central.com/report/coursera-cracked-chinese-market/

Specht, J., B. Egloff, and S.C. Schmukle. 2011. Stability and Change of Personality Across the Life Course: The Impact of Age and Major Life Events on Mean-level and Rank-order Stability of the Big Five. *Journal of Personality and Social Psychology* 101: 862-882.

Walsh, K. 2001. *Teacher Certification Reconsidered: Stumbling for Quality*. Baltimore: Abell Foundation.

World Bank. 1997. Expanding the Measure of Wealth: Indicators of Environmentally Sustainable Development. Environmentally Sustainable Development Studies and Monographs Series No. 17. Washington, D.C.

Yang, Y., B. Pan, and H. Song. 2014. Predicting hotel demand using destination marketing organization's WEB traffic data. *Journal of Travel Research*, 53(4): 433-447.

Yukl, G. 2008. How Leaders Influence Organizational Effectiveness. *The Leadership Quarterly* 19: 708-722.

Zaccaro, S.J., C. Kemp, and P. Bader. 2004. Leader traits and attributes. In Antonakis J., Cianciolo A.T., and Sternberg R.J. (Eds.), The Nature of Leadership, Thousand Oaks, CA: Sage, 101–124.

Zaccaro, S.J. 2007. Trait-based Perspectives of Leadership. *American Psychologist* 62(1): 6–16.