UNIVERSITAT DE BARCELONA
BUSINESS SCHOOL

MSc
Business Research

MASTER THESIS

Occupational Socio-Economic Status Returns to English Proficiency in Spain

MSc IN BUSINESS RESEARCH

University of Barcelona

Author: Junwei Fang

Director: Antonio Di Paolo

Barcelona, June 11, 2019

TABLE OF CONTENTS

Abstract

This paper estimates the occupational socio-economic status (ISEI-08) returns to English proficiency for Spanish workers with data from the Adult Education Survey. The estimations, either before or after correcting the selection bias, show that compared with respondents without English skills, those with at least elementary English proficiency share a higher ISEI-08 score, and such improvement continues to expand as they gain more English knowledge. Estimations separated by gender indicate no gender differences in the returns to English, although such conclusion should be taken with caution. Also, the results highlight the importance of education and suggest that female workers can benefit more from education than their male counterparts. This paper differs from similar studies by employing a new indicator and by contextualizing in non-English-speaking countries. It justifies the worldwide investment in English learning and calls for more attention on the gender differences in the returns to education.

Keywords: English proficiency, occupational socio-economic status (ISEI-08), employment

JEL Codes: I26, J24, M53, Z13

Resumen

El presente estudio estima el rendimiento del conocimiento del idioma inglés de los trabajadores españoles con estatus socio-económico ocupacional (ISEI-08) como indicador, basado en la EADA-2016. Los resultados muestran que, comparado con los trabajadores sin ningún conocimiento del inglés, los que tienen al menos un nivel elemental obtienen una nota de ISEI-08 más alta. Dicho incremento sigue progresando mientras mejoran su inglés. Las estimaciones separadas por sexo no encuentran diferencias entre los dos géneros. Los resultados también destacan la importancia de la educación, de la cual las mujeres se benefician más que los hombres. Este estudio se distingue de otros por ambientarse en un país no anglófono y por emplear un nuevo indicador. Los resultados justifican la inversión mundial en el aprendizaje del inglés y demandan más atención sobre la diferencia entre géneros con respecto al rendimiento de la educación.

Palabras Clave: conocimiento de inglés, estatus socio-económico ocupacional (ISEI-08), empleo

Códigos de Campos (JEL): I26, J24, M53, Z13

1. Introduction

The prevalence of English around the world has made it one of the most important languages in the labor market (Ginsburgh & Prieto-Rodriguez, 2013). It has even been included in the curriculum in many countries through macro-level national policies (Hamid & Nguyen, 2016; Portiño, 2018). Considering the enormous amount of investment of time and financial resources in English learning, it is then urgently important to calculate the returns to English knowledge. If the returns to mastering English do not balance the efforts invested, the necessity and correctness of certain personal decisions and even national policies may need to be better justified.

Numerous studies have confirmed that workers' English skills are positively related to their earnings (Chiswick & Miller, 1999; Shields & Price, 2002; Bleakley & Chin, 2004). These studies unanimously used earnings as the indicator to measure the returns to English knowledge. Albeit being a useful one, it might be beneficial to find other indicators so as to look at this problem from a different perspective. Apart from containing information related to earnings, such new indicators should also reflect the essential differences that exist among different types of occupations. According to Ganzeboom & Treiman (2003), "the division of labor is the kernel of social inequality", and occupational status measures can be derived from information on occupations. Such measures may combine information on occupational classifications together with education level, earnings, market conditions, and social rewards like approval or admiration. Obviously, occupational status measures can provide a comprehensive gauge of the bonus effect of English proficiency for workers.

Another common research line in the study of the returns to English knowledge in the labor market is to investigate whether gender differences exist (Ginsburgh & Prieto-Rodriguez, 2013; Wang et al., 2017), i.e., *ceteris paribus,* if two workers of opposite genders have a similar level of English proficiency, who would gain a better return? Again, in this type of studies, the economic returns are usually used as the criterion to reflect potential gender differences, whilst the gender differences may exist in more forms than just discrepancies in earnings. Occupational status measures have the potential to be an alternative indicator in this scenario for the same reasons aforementioned.

In order to enrich the studies on the returns to English knowledge and get better estimations, the objective of this paper is to measure the returns to English proficiency by using occupational status scales as the outcome, and to look for potential gender

differences in this relationship. Besides offering a distinct perspective by using a new indicator, i.e., occupational status scales, this study also differs from other studies by contextualizing itself in non-English-speaking countries. Many studies on this topic have centered on the earnings returns to immigrants in English-speaking host countries (Chiswick & Miller, 1999; Shields & Price, 2002; Bleakley & Chin, 2004). Although they seem to be a convenient case to study the returns to English proficiency, the immersion of the immigrants in the labor market in the host country is more than often different from that of the natives. It involves a more complicated mechanism since immigrants need to deal with more problems than just languages, e.g., cultural differences, when starting to work in the host country. By targeting at only non-English-speaking countries, this study could minimize the complexity of this relationship and thus reduce potential bias risks.

The rest of this paper is organized as follows: in the second section, related existing literatures are mentioned to better contextualize the objective of this study, which would lead to two hypotheses that this paper tries to validate; next, the third part would describe the sample and data used in this paper, together with some basic descriptive statistics; then, the empirical methods employed to test the hypotheses would be explained; after that, the results of the analysis are presented and discussed, leading to the conclusions of this paper; in the end, the academical contributions and managerial implications of this paper will be addressed along with some limitations and potential future research proposals.

2. Literature Review

2.1 Measuring the Returns to English Proficiency

The estimation of the returns to English knowledge (or more generally, the proficiency of any language) is an on-going study that has been carried out many times. They distinct from each other by contextualizing in a different country or area, e.g., the United States (Bleakley & Chin, 2004), India (Azam et al., 2013), South Africa (Casale & Posel, 2011), western Europe (Williams, 2011), China (Wang et al, 2017); or by employing a different set of empirical tools, e.g., instrumental variables estimation (Shields & Price, 2002), instrumental variables quantile regression (Ginsburgh & Prieto-Rodriguez, 2013), OLS regression (Fabo, 2017), multinomial probit model (Di Paolo & Tansel, 2019), etc. In general, they can be divided into two types based on the group of observations: those that study the returns to English for foreign immigrants in English-speaking countries (Chiswick & Miller, 1999; Shields & Price, 2002; Bleakley & Chin,

2004), and those that focus on the English skills of the natives in non-English-speaking countries (Ginsburgh & Prieto-Rodriguez, 2011; Di Paolo & Tansel, 2015; Di Paolo & Tansel, 2019). In order to get better estimations of the returns to English proficiency, however, the second type of studies might work better. First, as mentioned before, the mechanism behind the returns to English is more complex for foreign immigrants in English-speaking host countries since they are faced with issues that do not affect the natives, e.g., cultural differences. Second, as stated by Ginsburgh & Prieto-Rodriguez (2011), there is simply less market pressure to pay higher wages to immigrants that can speak the national language, which could distort the relationship of interest. They also discovered that in non-English-speaking EU countries, the earnings returns to English are usually higher than any other foreign languages. Di Paolo & Tansel (2015, 2019) carried out a similar research with Turkish national data. Their findings suggest that the earnings returns to English proficiency are present in both male and female workers. Overall, even though the second type of studies can get more exact estimations, these two types of studies are not well balanced since most studies are of the first type. This paper constitutes a further step in the second type of studies by contextualizing in non-English-speaking countries.

Another thing worth noticing is that these studies aforementioned unanimously chose to measure the returns to English proficiency from an economic perspective, i.e., earnings, which is a convenient and self-explanatory indicator. Some recent papers, however, have started to study this topic with other indicators, one of which being occupational status measures (Hamid & Nguyen, 2016; Pinilla-Portiño, 2018).

A classical sociological hypothesis believes that in social interactions, people tend to judge the social status of interaction partners by their occupations (Ganzeboom & Treiman, 2003). Reasonable or not, occupational titles are usually associated with a bundle of concepts: earnings, social position, education level, etc. The amount of implied information contained in occupations itself is surprisingly rich. Social researchers have noticed this phenomenon decades ago. Since then, they have created several important status measures based on the occupational classifications (Connelly et al., 2016). The construction of such measures varies one from another, but its core idea is similar: to make use of the information that occupational titles give away.

In the case of studying the returns to English proficiency, there are at least two reasons that make these occupation-based status measures have the potential to be an alternative indicator to earnings. First, these measures are closer to the actual returns

to English proficiency because they have a more complex construction. Unlike earnings, occupational status measures acknowledge the inherent horizontal (sectoral) and vertical (hierarchical) differences that exist among different occupations. They are usually a combination of elements related to a certain occupation. Depending on the researcher's needs (undoubtedly, such constructions are developed with strong theory bases), its construction could link education level, earnings, occupational classifications, employment situations, market conditions, and many more important pieces of information together (Ganzeboom & Treiman, 2003). Second, they provide more choices, hence more perspectives to view the same problem. There are many occupational status measures from regional scales to international standards (Connelly et al., 2016). Each one of them has its own strict principles of construction, which allows more flexibility at the time of deciding the angle of analysis. When trying to calculate the returns to English proficiency, these measures present a better reflection of the reality.

Nonetheless, using occupational status measures to study the returns to English is still in its beginning phase. Hamid & Nguyen (2016) and Pinilla-Portiño (2018) reported on the massive national investment in English learning in several Asian countries like Malaysia, Japan, Vietnam, and Bangladesh, arguing that such policies were made under the expectation that English will improve the employability of its citizens and enhance learners' social-economic development and social mobility. However, they did not provide further data support on such arguments. To continue their work, this paper intends to gauge the returns to English proficiency by employing occupational status measures in order to get more exact estimations.

2.2 Occupational Status Measures

To obtain occupational status measures from occupational information, stratification researchers usually follow two steps. First, establish occupational classifications that contain hundreds of categories (normally borrowed from census or other official classifications); second, recode such detailed classifications into status scales that have a more workable size and contain more sociological information (Ganzeboom & Treiman, 2003).

There are three widely used occupational status scales that may serve as a more proper indicator of the returns to English proficiency: occupational prestige measures, socio-economic status scales, and nominal class categories (Ganzeboom & Treiman, 2003; Connelly et al., 2016). All three have their own international standards coded on International Standard Classification of Occupations (ISCO), the most recognized

occupational classification system. Currently its latest version is ISCO-08, endorsed by International Labor Organization (ILO) in March, 2008. ISCO-08 consists of a four-digit hierarchical system that classifies all the occupations into 436 unit groups, which can be further aggregated into 130 minor groups, 43 sub-major groups, and finally, 10 major groups (ILO, 2016).

The National Opinion Research Center carried out a series of surveys since the 1940s that provided the original data to develop the occupational prestige measures. Despite of its constant version updates due to new occupational classifications and varied coding systems, the central idea of construction of this status scale basically remains untouched (Ganzeboom & Treiman, 2003). The prestige measures are generated from the popular evaluation of occupations according to their social standing, i.e., sensed prestige of each occupation (Nakao & Treas, 1994). It provides "a hierarchical ranking from the least to the most esteemed occupations according to average ratings" by surveyed respondents (Connelly et al., 2016). The most used occupational prestige measures in an international context is the Standard International Occupational Prestige Scale (SIOPS).

The most widely used international version of socio-economic status scales was constructed in 1992 and denominated International Socio-Economic Index of occupational status (ISEI) by a team led by Ganzeboom (Ganzeboom et al, 1992). It is a weighted sum of socio-economic characteristics of incumbents of each occupation, typically education and income, sometimes with adjustments for age (Ganzeboom, 2010; Connelly et al., 2016). The idea behind this construction is that education level influences earnings mainly through occupation, as shown in Figure-1. Occupational status is thus defined as the ability of occupations that turns educational qualifications into earnings. Hence, ISEI scores "are estimated as an optimal scaling of detailed occupation groups as a mediating variable" (Ganzeboom, 2010). Both ISEI and SIOPS are continuous scales, giving them huge advantages in statistical analysis. The latest version of ISEI is based on ISCO-08, i.e., ISEI-08, which is a two-digit system where a higher score means higher socio-economic status. It ranges from 10 (e.g., subsistence crop farmers) to 89 (e.g., generalist medical practitioners).

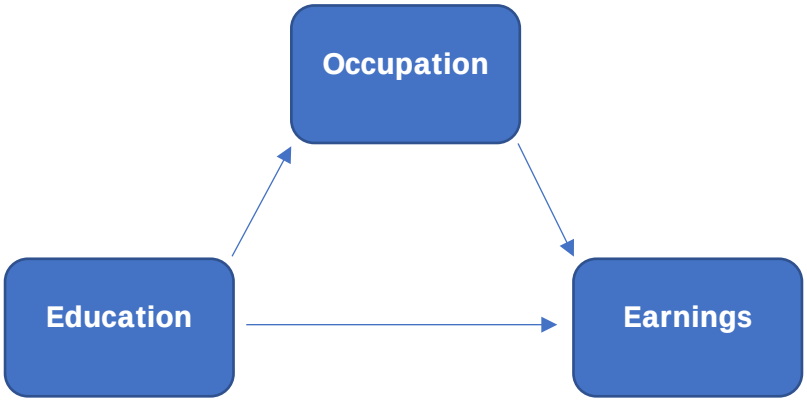

Figure-1 ISEI Construction Model

The Erikson-Goldthorpe-Portocarero (EGP) schema is considered by many the *de facto* international standard of nominal class categories (Erikson et al., 1979). According to its principles, employment relations decide the social class category of each individual (Erikson & Goldthorpe, 1992). It combines market situation (earnings and sector of work) with occupational situation (skill levels and supervisory status) (Ganzeboom & Treiman, 2003; Connelly et al., 2016). Individuals within the same social class category are believed to face similar life circumstances (Connelly et al., 2016). Obviously, EGP schema is a discrete scale.

These three occupational status measures represent different preferences at the time of statistical analysis: the choice between categorical and continuous approaches to occupational stratification. Such choice depends on the objectives of each study. Afterall, they usually render similar results (Ganzeboom & Treiman, 2003). However, Coxon & Jones (1978) questioned the consistency of SIOPS due to the cognitive issues caused by letting individuals to rank occupations. Their experiments showed that even the same observation could give very different rankings under different context, and that people tend to give their own occupation a higher score, the so-called “occupational egoism”. As for the EGP schema, it underestimates the power of occupational hierarchy in social stratification (Bergman & Joye, 2005). Also, its multiple categories (could easily range from 7 to 11) stand a challenge for multivariate statistical analysis (Connelly et al., 2016). So, in this study, ISEI-08 will be used as the indicator of the returns to English proficiency.

As mentioned before, the construction of ISEI-08 considers earnings, which means that when earnings are used as an indicator to measure the returns to English proficiency, the results are probably a partial reflection of what ISEI-08 would have shown. Since multiple studies have confirmed that English proficiency is positively related

to earnings, plus the statements aforementioned by Hamid & Nguyen (2016) and Pinilla-Portiño (2018), it is not too imprudent to reach the following hypothesis:

Hypothesis 1: An increase in English proficiency is positively related to workers' occupational socio-economic status.

2.3 Gender Differences in the Returns to English Proficiency

Researchers studying the economic returns to English proficiency have also been looking for the existence of gender differences. However, they have not reached to an agreement. Ginsburgh & Prieto-Rodriguez (2013) insisted that English knowledge can benefit both male and female workers, but to a different extent in different countries. In France, Italy, and Spain, gender differences are present. Female workers earn less than their male counterparts even though they have a similar level of English skills. While in Germany, Finland, and Denmark, there is no significant difference in earnings for women and men with similar English proficiency. On the other hand, Wang et al. (2017) got the opposite results. Their paper claimed that in China, based on national survey data and several different methods of estimations, the economic returns to English knowledge are higher for women than for men. Although current studies do not render one unified conclusion, there is one explanation for the potential effect of gender differences. Mora & Davila (1998) suggested that the differences in economic returns to English proficiency came from the occupational choices by men and women. They believe that men and women have different preferred types of occupations (either due to personal decisions or society conventions), resulting in that they are "occupational crowded" in certain occupations. When either gender is more present in certain occupations, the economic returns to English proficiency for this gender would decrease accordingly in these occupations owing to excess labor supply of such gender.

Similarly, these studies focusing on the gender differences also tend to use earnings as the indicator of the returns to English proficiency. Based on the same reasonings aforementioned, this study intends to employ ISEI-08 as the indicator. Since no agreement has been reached yet on this subject, and considering that the explanation by Mora & Davila (1998) is gender-neutral, this paper proposes the following second hypothesis that would be tested later:

Hypothesis 2: Occupational socio-economic status returns to English knowledge are different for female and male workers with the same level of English.

3. Data Description

This paper analyzes the returns to English in Spain, where English is one of the most used workplace languages. Its increasingly active labor market after the financial crises makes it an interesting case to study. The analysis borrows data from the Spanish version of the Adult Education Survey[1] (AES). It is a recurrent survey carried out by the National Statistics Institute of Spain to gather information on lifelong learning phenomenon. The AES has been conducted three times so far (in 2007, 2011, and 2016 respectively). Due to the minor discrepancies that exist in the survey methodologies over the years, this study focuses on the latest survey data (AES-2016) only.

AES-2016 was conducted in the Spanish territory to obtain information on adult (between 18 and 64 years old) learning activities carried out during the 12 months prior to the interview, which embraces a wide range of information that could be employed for the purpose of this study. More specifically, among others, AES-2016 provides information on individual and family characteristics like gender, age, level of education, the number of family members of different age groups, labor market situation, and sociodemographic information, etc. The most important information that AES-2016 contains, however, is that this survey includes a section where it records the knowledge of different languages of each individual. Respondents were required to report at most two mother tongue languages and up to seven foreign languages that they can speak. Also, when applicable, they were asked to indicate the proficiency levels of the two foreign languages that they know best. Language proficiency is divided in four levels[2] (below are the exact words employed in the questionnaire):

1. I only understand and can use a few words and phrases.
2. I can understand and use the most common everyday expressions. I use the language in relation to familiar things and situations.
3. I can understand the essential of clear language and produce simple text. I can describe experiences and events and communicate fairly fluently.
4. I can understand a wide range of demanding texts and use the language flexibly. I master the language almost completely.

[1] Survey name in Spanish: "Encuesta sobre la Participación de la Población Adulta en las Actividades de Apredizaje", located at: http://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176759&menu=resultados&secc=1254736194656&idp=1254735573113.

[2] Respondents without any knowledge of English are coded "0" during the analysis.

The entire microdata of this survey contains 23,019 observations. In order to minimize possible biases, 1,746 observations that do not have Spanish as at least one of their mother tongue languages are excluded. 155 respondents with unknown level of English proficiency are then removed since it constitutes an important information for this study. Also, those (3,379 cases) that were studying, retired, permanently disabled, or with unknown job-related information at the time of the interview are deleted. Only respondents that were working (including 8,994 employees, 1,296 self-employed workers, 624 salaried employers, and 57 salaried family members in family businesses), unemployed, or were involved in household works are considered as valid observations. In AES-2016, the respondents' occupations (if applicable) were coded using ISCO-08. However, only the first two digits of ISCO-08 were used, i.e., major and sub-major groups of occupations. Accordingly, this paper transforms ISCO-08 into ISEI-08 scores (Ganzeboom, 2010) basing on these two digits. The loss of information due to the missing of the last two digits of ISCO-08 results in corresponding ISEI-08 scores with a range from 10 to 69, while the whole range of ISEI-08 is between 10 and 89. However, since the major and sub-major groups of ISCO-08 already reflect the general structure of occupational classifications, this loss of information is considered affordable and will not change the landscape of the results. The reason why respondents that were unemployed or were involved in household works are also retained for the analysis is to control for potential selection bias, which will be explained in more details later. Another 2,121 observations with other missing values that would be used as control variables for the analysis are omitted, too. In the end, 15,618 observations are used for this study.

In the final sample, 46.07% are male individuals. 9,092 of them were working part-timely and 1,927 full-timely. 21.56% of the observations were unemployed at the time of the survey while another 7.88% were involved in unpaid household works. The age distributions between male and female are rather similar. Overall, the average age was 44.9 years old with a total range from 18 to 64. More detailed descriptive statistics can be found in the Appendix, with separate summaries of variables based on respondents' labor market situation (working or not working). These summaries show that among working respondents, 58.8% have no knowledge of English; while this figure surges to 73.28% for those that were not working. 24.22% working respondents possess intermediate or higher levels of English, compared with only 13.92% for non-working respondents. On the other hand, respondents that had completed only secondary education or less represent 30.61% among those that were working at the time of the survey, and this share jumps to 41.7% among those that were not working. Moreover,

30.93% working respondents had a university degree or higher, compared with only 12.05% for non-working respondents. These numbers suggest that English proficiency and education level constitute key competences for entering the labor market.

As for the regressor of interest (ISEI-08), Figure-2 shows the box plots of ISEI-08 scores by respondents' levels of English proficiency. Obviously, there is a positive relationship between these two variables. Respondents with advanced English proficiency in average share 14.28 units ISEI scores higher than those without any knowledge of English. It is a rough confirmation of the first hypothesis. After controlling for related variables like age, gender, level of education, etc., the estimations of the *ceteris paribus* returns to English proficiency will be obtained.

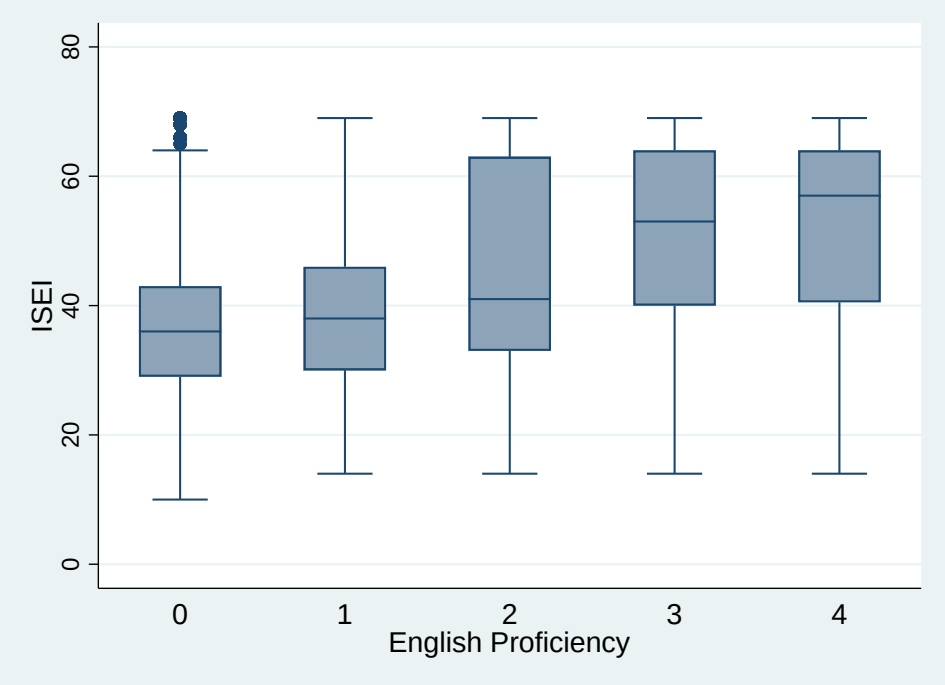

Figure-2 Box plots of ISEI-08 scores by English proficiency levels: no knowledge of English = 0, very basic proficiency = 1, elementary proficiency = 2, intermediate proficiency = 3, advanced proficiency or mother tongue = 4

Figure-3 shows the share of individuals with different levels of English proficiency categorized by their labor market situation. The histograms of workers, part-time or full-time, are surprisingly similar: over one fifth (24.22% and 24.24% respectively) had an intermediate or advanced level of English, while less than 60% had no knowledge of English. However, in the case of unemployed respondents, 69.27% reported having no knowledge of English and only 16.54% of them could speak English properly (intermediate or higher level). As for those involved in household unpaid works, the

share of individuals without any knowledge of English reached 84.24%. Less than one tenth (7.47%) considered themselves with a proper level of English proficiency.

Since ISEI-08 is based on occupational classifications, Figure-3 does not necessarily facilitate the analysis of the relationship of interest due to the inclusion of unemployed individuals and respondents dedicated to household works. However, the distribution of observations of different English proficiencies does help to indicate that certain selection bias problem may exist in the microdata. Maybe (it could be the other way around, but either way it would produce a selection bias) people with less knowledge of English had a harder time entering the labor market so they were unemployed or decided to turn to unpaid household works. Hypothetically, if they did ever enter the labor market, they would be occupying occupations that have relatively lower ISEI-08 scores. The analysis of the relationship of interest without considering this self-selection into the labor market would then underestimate the effect of English proficiency on ISEI-08 scores since those with lower English skills that would have occupied a low occupational socio-economic status are omitted because they were not in the labor market in the first place. Test for the existence of this issue and corresponding solutions will be further discussed in the next section.

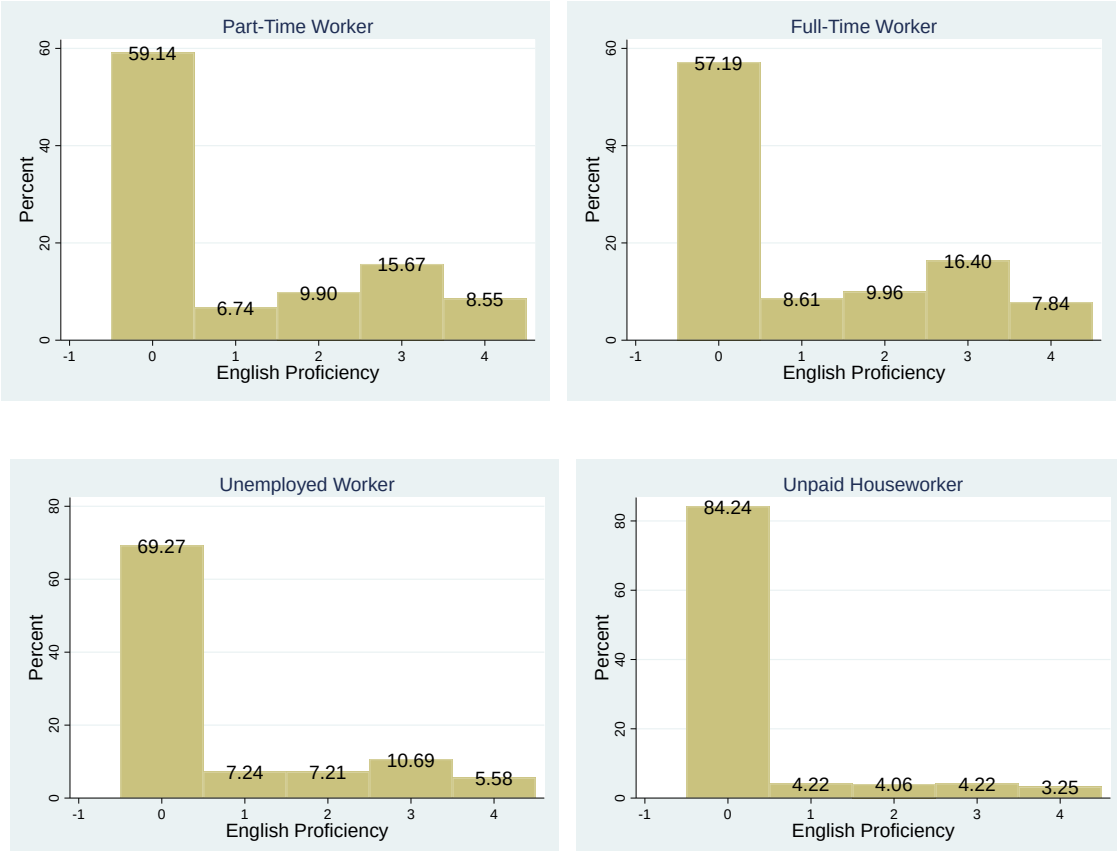

Figure-3 Share of respondents' English skills by labor market categories: no knowledge of English = 0, very basic proficiency = 1, elementary proficiency = 2, intermediate proficiency = 3, advanced proficiency or mother tongue = 4

4. Empirical Methodology

In order to test the effect of English proficiency on ISEI-08 scores, the first attempt is an OLS regression that uses ISEI-08 (Y_i) as the dependent variable, with English proficiency and a set of control factors as the independent variables, i.e.:

$$Y_i = \alpha + \beta' X_i + \delta_1 I(EN_i = 1) + \delta_2 I(EN_i = 2) + \delta_3 I(EN_i = 3) + \delta_4 I(EN_i = 4) + \varepsilon_i \qquad (1)$$

In this case, the dummy variables EN_i indicates the level of English proficiency (very basic = 1, elementary = 2, intermediate = 3, advanced or mother tongue = 4) of each respondent, with no knowledge of English ($EN_i = 0$) as the base group. Their coefficients δ show the conditional changes in occupational socio-economic status relative to the base group, i.e., respondents with no knowledge of English. A set of other factors that may influence workers' occupational socio-economic status are contained in X_i, ranging from individual and family characteristics like gender, age and its square, education level, highest education obtained by parents, whether or not living with a partner, number of co-living family members of different age groups (0 to 13 years old, 14 to 24 years old, 25 years old or more), to sociodemographic factors like the size of the municipality and in which autonomous communities the survey was carried out. They are added stepwise for a clearer understanding of the effects of these control variables.

Apart from common issues for regression analysis like heteroscedasticity, which would be tested and corrected during the analysis, one potential issue with this equation is that it only includes people that were working (thus having a corresponding ISEI-08 score) at the time of the survey, while these individuals are not strictly a random sample of the entire labor force, as mentioned in the third section. Some latent factors may affect the labor force participation, which would lead to a selection bias for the relationship to be studied, meaning that the estimates obtained for the subsample of employed individuals are inconsistent and not representative for the whole population. Heckman (1976) proposed a two-step correction[3] for this type of sample selection issues: a first-step selection equation that determines the probability that the dependent variable would be observed, proceeded by a second equation of the liner model of

[3] This paper employees one-step estimation using maximum likelihood, which is more efficient and easier to implement in Stata, the software used in this study.

interest already corrected for selection bias. In order to use this method, at least one exclusion restriction is needed, which would be proposed later. The selectivity-corrected estimations will be compared with former OLS regressions to check whether selection bias constitutes a serious problem in this study.

The last part of the empirical analysis consists in investigating the effect of gender in the relationship of interest. Here, the basic idea is similar to equation (1), except that the sample will be divided in two groups by gender and estimated respectively. The comparison of the results would show if there are any gender differences concerning the returns to English proficiency.

5. Results and Discussions

5.1 OLS Regression Results

When implementing the OLS regressions, factors other than the level of English proficiency that may affect workers' ISEI-08 scores are added as control variables by batch. This way, it is easier to observe the evolution of the returns to English proficiency when different factors are maintained constant, i.e., the *ceteris paribus* effect of English proficiency. In all the regressions, robust estimations are used to correct heteroscedasticity.

In the first column, only the level of English proficiency is used as the independent variable. It shows that when other factors are uncontrolled, i.e., the unconditional difference, relative to workers without any knowledge of English (base category), those with basic English skills enjoy a 2.31-unit increase in ISEI-08 scores, while those with elementary English skills get 7.48 units more. Workers with intermediate (a 13.25-unit increase) and advanced (a 14.28-unit increase) English proficiencies do not seem to have a very different occupational socio-economic status (ANOVA test confirms that their coefficients are not significantly different from each other at a 0.1 level). It can be argued that an intermediate level of English is already enough for most work scenarios. A further improvement in ISEI-08 scores is more influenced by other factors, e.g., level of education. For the second regression, some very basic individual variables are added: gender, age and its square. These control variables do not seem to interfere too much the relationship of interest. The coefficient of the dummy for gender is not significant, while age has a quadratic effect on ISEI-08 scores, meaning that overall occupational socio-economic status tends to increase with age but at a decreasing rate between 33 and 64 (the upper range of the variable age) years old.

In the third and fourth column, two dummy variables are added one after the other: the highest education obtained by parents and that by the respondents. Relative to respondents whose parents had obtained only primary education or less, those whose parents had at least finished tertiary education can get much higher (more than 6 units on average) ISEI-08 scores. This could be explained by the fact that parents with higher education levels may have better social networks, so that they can help their children to obtain better education and better job opportunities. After introducing the respondents' education level, the adjusted R-squared shows an obvious improvement (from 0.15 to 0.2 in the third regression and to 0.45 in the fourth regression). This demonstrates that individual's education level takes account for an important proportion of the variance in occupational socio-economic status. The results show that workers with a university or higher degree share a much higher ISEI-08 score (29.03 units more) than those with only primary education or less. On the other hand, while the inclusion of the education level of parents affects moderately the coefficients of English proficiency, the respondents' education level reduces substantially the coefficients of the English proficiency dummies. This phenomenon can be explained as that parents with higher education levels may pay more attention to the education of their children, while workers with better education background are usually more welcomed in the labor market, hence reaching a higher occupational socio-economic status than their peers with a lower level of education. Moreover, people with better education may acquire better English proficiency since English learning is also a part of education. Nevertheless, even after controlling the most influential factors, i.e., respondents' education level, the general pattern remains the same. Workers with more English knowledge still tend to have a higher ISEI-08 score, with a difference of around 3 units among workers without any knowledge of English and those with at least intermediate English proficiency. Also, gender begins to matter in this relationship. Female workers in general get 2.63 units less than their male counterparts.

As for the last two columns, two territorial variables (dummies for municipality size and for the autonomous community of residence) and four household characteristics (whether or not living with a partner and the number of co-living family members of three different age groups) are added. Despite of the many control variables added in these two regressions, the coefficients of existing independent variables (including the English proficiency dummies) do not seem to be affected too much. Even the adjusted R-squared remains at a similar level as in the fourth regression. However, these two regressions do present several interesting results. Living in a small municipality (less

than 20.000 residents) is negatively related to occupational socio-economic status, since there may be less high-quality labor offers than larger municipalities. Respondents from Baleares and Catalunya in average shared at least 1-unit higher ISEI-08 score than those in Andalusia. Living with a partner seems to have a positive effect on their occupational socio-economic status, too. Finally, number of co-living family members with more than 25 years old lowers slightly ISEI-08 scores at a rate of 0.58 unit.

Overall, the results confirm the first hypothesis that higher levels of English proficiency are positively related to a higher occupational socio-economic status. When other factors are controlled, workers with even just elementary English proficiency show an improvement in their ISEI-08 scores by about 1.75 units relative to those without English knowledge. Such improvement continues to grow as they gain more knowledge of English. On the other hand, respondents' education levels can explain over one fifth of the variations in occupational socio-economic status, which is in line with many other related studies. Respondents with a university degree or higher in average get 28.64 units more in their ISEI-08 scores relative to those with only primary education or less. Being significant even after such influential factors are controlled shows that English proficiency can indeed bring improvements to workers' occupational socio-economics status.

Table-1 OLS Regressions Results

Dependent Variable	(1) ISEI	(2) ISEI	(3) ISEI	(4) ISEI	(5) ISEI	(6) ISEI
Constant	37.50***	21.69***	13.83***	16.49***	16.99***	21.95***
	(0.18)	(1.99)	(1.94)	(1.74)	(1.76)	(2.02)
No English skills			*Reference Category*			
Basic skills	2.31***	2.59***	2.17***	0.28	0.28	0.28
	(0.53)	(0.53)	(0.53)	(0.44)	(0.45)	(0.45)
Elementary skills	7.48***	7.86***	6.69***	1.84***	1.77***	1.75***
	(0.48)	(0.48)	(0.47)	(0.40)	(0.41)	(0.40)
Intermediate skills	13.25***	13.88***	11.18***	2.99***	2.84***	2.84***
	(0.39)	(0.39)	(0.41)	(0.37)	(0.37)	(0.37)
Advanced skills	14.28***	14.95***	11.51***	3.47***	3.39***	3.41***
	(0.53)	(0.53)	(0.54)	(0.45)	(0.45)	(0.45)
Man			*Reference Category*			
Woman		-0.11	-0.26	-2.63***	-2.69***	-2.66***
		(0.27)	(0.27)	(0.22)	(0.22)	(0.22)
Age		0.66***	0.86***	0.30***	0.29***	0.16*
		(0.10)	(0.09)	(0.08)	(0.08)	(0.08)
Square of age		-0.01***	-0.01***	-0.00***	-0.00***	-0.00
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Highest education obtained by parents: primary education or less			*Reference Category*			
Secondary education			2.98***	1.14***	1.08***	1.06***
			(0.35)	(0.29)	(0.29)	(0.29)
Tertiary education			6.53***	1.97***	1.85***	1.84***
			(0.46)	(0.39)	(0.39)	(0.39)
University or higher			10.65***	3.69***	3.48***	3.48***
			(0.43)	(0.38)	(0.38)	(0.38)
Number of observations	11019	11019	11019	11019	11019	11019
Adjusted R-squared	0.14	0.15	0.20	0.45	0.45	0.45

Note: ***, **, * denote significance at 1, 5, and 10 percent level respectively; robust standard errors in parentheses

Table-1 (continued) OLS Regressions Results

Dependent Variable	(1) ISEI	(2) ISEI	(3) ISEI	(4) ISEI	(5) ISEI	(6) ISEI
Highest education obtained by respondent: primary education or less				*Reference Category*		
Secondary education				6.40***	6.37***	6.26***
				(0.64)	(0.64)	(0.64)
Tertiary education				13.57***	13.40***	13.25***
				(0.64)	(0.65)	(0.65)
University or higher				29.03***	28.82***	28.64***
				(0.68)	(0.69)	(0.69)
Capital of province or with more than 100,000 residents				*Reference Category*		
Municipality with between 50,001 and 100,000 residents					-0.02	-0.03
					(0.42)	(0.42)
Municipality with between 20,001 and 50,000 residents					-0.20	-0.24
					(0.34)	(0.34)
Municipality with between 10,001 and 20,000 residents					-0.90**	-0.95***
					(0.37)	(0.36)
Municipality with 10,000 or less residents					-1.46***	-1.45***
					(0.33)	(0.33)
Andalucía				*Reference Category*		
Aragón					0.83	0.76
					(0.58)	(0.58)
Asturias					0.86	0.85
					(0.67)	(0.67)
Baleares					1.44**	1.34**
					(0.65)	(0.65)
Canarias					-1.06*	-1.01
					(0.64)	(0.64)
Number of observations	11019	11019	11019	11019	11019	11019
Adjusted R-squared	0.14	0.15	0.20	0.45	0.45	0.45

Note: ***, **, * denote significance at 1, 5, and 10 percent level respectively; robust standard errors in parentheses

Table-1 (continued) OLS Regressions Results

Dependent Variable	(1) ISEI	(2) ISEI	(3) ISEI	(4) ISEI	(5) ISEI	(6) ISEI
Cantabria					-0.62	-0.57
					(0.66)	(0.66)
Castilla y León					0.80	0.73
					(0.55)	(0.55)
Castilla-La Mancha					1.17**	1.09*
					(0.57)	(0.57)
Cataluña					1.59***	1.51***
					(0.47)	(0.47)
Valencia					0.25	0.18
					(0.51)	(0.52)
Extremadura					0.26	0.21
					(0.60)	(0.60)
Galicia					0.04	0.09
					(0.56)	(0.57)
Madrid					0.78	0.78
					(0.48)	(0.48)
Murcia					-0.62	-0.69
					(0.62)	(0.61)
Navarra					-0.46	-0.48
					(0.66)	(0.66)
País Vasco					0.43	0.40
					(0.57)	(0.57)
La Rioja					-0.11	-0.21
					(0.65)	(0.65)
Ceuta					-1.25	-1.06
					(1.13)	(1.12)
Melilla					-0.22	-0.14
					(1.41)	(1.42)
Number of observations	11019	11019	11019	11019	11019	11019
Adjusted R-squared	0.14	0.15	0.20	0.45	0.45	0.45

Note: ***, **, * denote significance at 1, 5, and 10 percent level respectively; robust standard errors in parentheses

Table-1 (continued) OLS Regressions Results

Dependent Variable	(1) ISEI	(2) ISEI	(3) ISEI	(4) ISEI	(5) ISEI	(6) ISEI
Living with a partner			*Reference Category*			
Does not live with a partner						-1.13***
						(0.29)
Number of co-living family members over 25 years old						-0.58***
						(0.15)
Number of co-living family members between 14 and 24 years old						-0.17
						(0.16)
Number of co-living family members between 0 and 13 years old						-0.11
						(0.15)
Number of observations	11019	11019	11019	11019	11019	11019
Adjusted R-squared	0.14	0.15	0.20	0.45	0.45	0.45

Note: ***, **, * denote significance at 1, 5, and 10 percent level respectively; robust standard errors in parentheses

5.2 Correction for Self-Selectivity

Selection bias constitutes a common problem for analysis where the variable of interest (in this case, the ISEI-08 scores) can only be observed when certain latent conditions are fulfilled (e.g., only those respondents that were working at the time of the survey would get a corresponding ISEI-08 score). To correct potential selection bias by Heckman's method, exclusion restrictions are needed. For the exclusion restrictions, this study proposes to use the percentage of co-living family members between 0 and 13 years old and the percentage of co-living family members between 14 and 24 years old. To be a suitable exclusion restriction for the Heckman correction method, such restriction has to fulfill two conditions at the same time. First, it cannot be related to the final outcome of interest. Here, the numbers of these two groups of co-living family members are added as control variables, while their percentages are used as exclusion restrictions. Second, such restriction has to be a factor that influences the probability of whether the dependent variable of interest is observed. When respondents co-live with several family members that are less than 25 or even 14 years old, there may be a higher probability that they will need to make a choice between working and taking care of such family members. Based on this reasoning, the two chosen identifying variables are negatively related to labor market participation, which would in the end decide whether the variable of interest is observed or not. Other studies have also shown a similar strategy (Di Paolo & Tansel, 2019). The results are presented in Table-2. The first column is the same full OLS regression (6) from Table-1, copied here for comparison. The second column is the main equation of interest corrected for selection bias, while the last column is the employment equation that shows the probability of entering the labor market.

Table-2 Sample Selection Correction

Dependent Variable	(1) ISEI (OLS)	(2) ISEI (Main)	(3) ISEI (Select)
Constant	21.95***	36.62***	-1.72***
	(2.02)	(2.26)	(0.19)
No English Skills		*Reference Category*	
Basic skills	0.28	0.12	0.05
	(0.45)	(0.48)	(0.05)
Elementary skills	1.75***	1.39***	0.10**
	(0.40)	(0.43)	(0.04)
Intermediate skills	2.84***	2.91***	-0.03
	(0.37)	(0.39)	(0.04)
Advanced skills	3.41***	3.39***	-0.03
	(0.45)	(0.48)	(0.05)
Man		*Reference Category*	
Woman	-2.66***	-0.37	-0.60***
	(0.22)	(0.25)	(0.02)
Age	0.16*	-0.27***	0.09***
	(0.08)	(0.09)	(0.01)
Square of age	-0.00	0.00***	-0.00***
	(0.00)	(0.00)	(0.00)
Highest education obtained by parents: primary education or less		*Reference Category*	
Secondary education	1.06***	0.81***	0.08***
	(0.29)	(0.31)	(0.03)
Tertiary education	1.84***	1.84***	0.01
	(0.39)	(0.41)	(0.04)
University or higher	3.48***	3.29***	0.05
	(0.38)	(0.40)	(0.04)
Highest education obtained by respondent: primary education or less		*Reference Category*	
Secondary education	6.26***	4.58***	0.29***
	(0.64)	(0.75)	(0.06)
Tertiary education	13.25***	9.66***	0.78***
	(0.65)	(0.76)	(0.06)
University or higher	28.64***	23.61***	1.24***
	(0.69)	(0.81)	(0.07)
Capital of province or with more than 100,000 residents		*Reference Category*	
rho		-0.70***	
		(0.02)	
lnsigma		2.55***	
		(0.01)	
N	11019	15618	15618
Adjusted R-squared	0.45		

Note: ***, **, * denote significance at 1, 5, and 10 percent level respectively; robust standard errors in parentheses

Table-2 (continued) Sample Selection Correction

Dependent Variable	(1) ISEI (OLS)	(2) ISEI (Main)	(3) ISEI (Select)
Municipality with between 50,001 and 100,000 residents	-0.03	-0.06	0.01
	(0.42)	(0.45)	(0.04)
Municipality with between 20,001 and 50,000 residents	-0.24	-0.47	0.04
	(0.34)	(0.36)	(0.03)
Municipality with between 10,001 and 20,000 residents	-0.95***	-1.23***	0.06
	(0.36)	(0.39)	(0.04)
Municipality with 10,000 or less residents	-1.45***	-1.68***	0.05
	(0.33)	(0.35)	(0.03)
Andalucía		*Reference Category*	
Aragón	0.76	-1.03*	0.41***
	(0.58)	(0.62)	(0.06)
Asturias	0.85	0.27	0.11*
	(0.67)	(0.72)	(0.06)
Baleares	1.34**	-0.50	0.40***
	(0.65)	(0.70)	(0.07)
Canarias	-1.01	-1.09	0.03
	(0.64)	(0.69)	(0.06)
Cantabria	-0.57	-1.36*	0.15**
	(0.66)	(0.71)	(0.06)
Castilla y León	0.73	-0.57	0.32***
	(0.55)	(0.59)	(0.06)
Castilla-La Mancha	1.09*	0.34	0.15***
	(0.57)	(0.61)	(0.05)
Cataluña	1.51***	-0.46	0.44***
	(0.47)	(0.51)	(0.05)
Valencia	0.18	-0.66	0.19***
	(0.52)	(0.55)	(0.05)
Extremadura	0.21	-0.06	0.05
	(0.60)	(0.64)	(0.06)
Galicia	0.09	-1.17*	0.27***
	(0.57)	(0.61)	(0.06)
Madrid	0.78	-1.08**	0.44***
	(0.48)	(0.52)	(0.05)
Murcia	-0.69	-1.97***	0.27***
	(0.61)	(0.66)	(0.06)
Navarra	-0.48	-2.39***	0.46***
	(0.66)	(0.71)	(0.07)
rho		-0.70***	
		(0.02)	
lnsigma		2.55***	
		(0.01)	
N	11019	15618	15618
Adjusted R-squared	0.45		

Note: ***, **, * denote significance at 1, 5, and 10 percent level respectively; robust standard errors in parentheses

Table-2 (continued) Sample Selection Correction

Dependent Variable	(1) ISEI (OLS)	(2) ISEI (Main)	(3) ISEI (Select)
País Vasco	0.40	-1.32**	0.42***
	(0.57)	(0.61)	(0.06)
La Rioja	-0.21	-2.05***	0.44***
	(0.65)	(0.70)	(0.07)
Ceuta	-1.06	-0.89	-0.05
	(1.12)	(1.25)	(0.11)
Melilla	-0.14	-0.29	0.09
	(1.42)	(1.52)	(0.12)
Living with a partner		*Reference Category*	
Does not live with a partner	-1.13***	-0.14	-0.21***
	(0.29)	(0.32)	(0.03)
Number of co-living family members over 25 years old	-0.58***	-0.27*	-0.05***
	(0.15)	(0.16)	(0.02)
Number of co-living family members between 14 and 24 years old	-0.17	-0.20	-0.11**
	(0.16)	(0.17)	(0.04)
Number of co-living family members between 0 and 13 years old	-0.11	-0.04	-0.13**
	(0.15)	(0.16)	(0.06)
Share of co-living family members between 0 and 13 years old			0.58**
			(0.24)
Share of co-living family members between 14 and 24 years old			0.49***
			(0.17)
rho		-0.70***	
		(0.02)	
lnsigma		2.55***	
		(0.01)	
N	11019	15618	15618
Adjusted R-squared	0.45		

Note: ***, **, * denote significance at 1, 5, and 10 percent level respectively; robust standard errors in parentheses

The results in Table-2 have confirmed the existence of selection bias in the sample. The value of rho, i.e., the correlation between the error term of the ISEI equation and the error term of the latent equation for self-selection into the labor market, is non-zero and significant at a 1% level. An extra joint F-test of the significance of the two exclusion restrictions in the selection equation also proves that the two identifying variables used are suitable for this analysis at a 1% level. Moreover, the employment equation shows that there are many factors that can affect the probability of entering the labor market. The average marginal effects of the coefficients in the last column indicate the conditional probability of self-selection into the labor market, e.g., compared with respondents with primary education or less, the probability of being working is raised by 10.46 percentage points with a secondary-education degree, by 27.09

percentage points with a tertiary-education degree, and by 38.54 percentage points with a university degree or higher. Relative to those living in Andalusia, such probability is increased by slightly over 10 percentage points for respondent residing in Aragon, Baleares, Catalunya, Madrid, Navarra, Basque Country, and La Rioja. Another factor that affects this probability by more than 10 percentage points is gender. Being a female reduces the probability of entering the labor market by 17.78 percentage points. Based on these figures, it is important to correct the selection bias in order to get better estimations of the returns to English proficiency.

Compared with the full OLS regression estimations without selection bias correction, the coefficients for English proficiency levels do not experience a remarkable change. Relative to those without English skills, workers with only very basic English proficiency still show no significant relation with the dependent variable, while the coefficients for those with elementary and advanced levels go through a minor diminution. The effect of intermediate English proficiency, on the other hand, has strengthened slightly, from 2.84 to 2.91 units. Still, the overall effect of English proficiency remains at a similar scale. However, the influence of individuals' education level shows a relatively big drop. In the case of workers with a university degree or higher, they now only get 23.61 units more than those with primary education or less, while this difference is 28.64 before correction. This is a sign that for this sample, a standard OLS regression analysis tends to overestimate the effect of education on occupational socio-economic status. This time, gender and whether living with a partner do not seem to affect significantly the relationship. Age and its square, however, have changed both their values and signs. ISEI-08 scores seem to drop very moderately as the age increases. The downside of living in a few-populated municipality is moderately amplified. Compared with those residing in the capital of province or a municipality with more than 100,000 residents, respondents from a municipality with less than 20,000 residents face at least a 1.23-unit drop in ISEI-08 scores.

5.3 Separate Estimations by Gender

Many studies have tried to find gender differences in the returns to English proficiency in the labor market (Ginsburgh & Prieto-Rodriguez, 2013; Wang et al., 2017). Even though the conclusions vary across studies as mentioned before, most of them do indicate that gender differences exist in terms of the earnings returns to English skills. Obviously, such discrepancies may also appear in occupational socio-economic status between workers of different genders. In order to understand how the ISEI-08 returns

to English proficiency differ by gender, this section carries out separate estimations by gender. Although previous sections have mentioned the problem of selection bias in this sample, the effect of English proficiency do not experience many variations before and after controlling the self-selection into the labor market as demonstrated. So, in this part, simple OLS regressions will be performed[4] and discussed. The results are displayed in Table-3, where the first column shows the estimations from the male sample and the second from the female sample.

According to the results, the constant for female respondents is 3.47 units less than male respondents and both statistically significant, which suggests that *ceteris paribus*, female workers have a relatively lower occupational socio-economic status than their male counterparts. In both cases, the coefficients of the English dummies are only significant from elementary levels. Workers of either gender get higher ISEI-08 scores when they have at least elementary English skills relative to those without English knowledge, but at a similar scale for both male and female respondent. A cross-model joint test confirms that the corresponding coefficients of the English dummies obtained from male and female sample are not significantly different from each other, which disagrees with the second hypothesis, i.e., ISEI-08 returns to English knowledge are the same for female and male workers with the same level of English. Interestingly enough, the returns to individual's education level does show a gender difference, and such difference enlarges as the respondents get higher education. Compared to male workers with only primary education or less, male workers with secondary, tertiary, university or higher education show an improvement in ISEI-08 scores by 4.66, 10.58, and 24.12 units, respectively. Meanwhile, for female workers, these figures climb to 7.34, 15.94, and 32.54 units, respectively. A joint test also shows that these two sets of coefficients are different from each other at a 5% level. It indicates that female workers may benefit more from obtaining more education than their male counterparts. The effect of the highest education level obtained by parents does not seem to show a similar gender difference.

[4] The two restriction instruments have low precision when used to estimate the sample by gender.

Table-3 Separate Estimations by Gender

Dependent Variable	(1) ISEI-M (OLS)	(2) ISEI-F (OLS)
Constant	22.35***	18.88***
	(2.88)	(2.78)
No English skills	*Reference Category*	
Basic skills	0.92	-0.24
	(0.63)	(0.62)
Elementary skills	1.72***	1.67***
	(0.59)	(0.55)
Intermediate skills	3.48***	2.32***
	(0.53)	(0.51)
Advanced skills	3.31***	3.76***
	(0.66)	(0.61)
Age	0.28**	0.03
	(0.12)	(0.12)
Square of age	-0.00	0.00
	(0.00)	(0.00)
Highest education obtained by parents: primary education or less	*Reference Category*	
Secondary education	0.89**	1.31***
	(0.40)	(0.42)
Tertiary education	1.16**	2.56***
	(0.55)	(0.55)
University or higher	3.26***	3.69***
	(0.55)	(0.53)
Highest education obtained by respondent: primary education or less	*Reference Category*	
Secondary education	4.66***	7.34***
	(1.03)	(0.66)
Tertiary education	10.58***	15.94***
	(1.05)	(0.67)
University or higher	24.12***	32.54***
	(1.11)	(0.73)
Capital of province or with more than 100,000 residents	*Reference Category*	
Municipality with between 50,001 and 100,000 residents	-0.61	0.71
	(0.58)	(0.61)
Municipality with between 20,001 and 50,000 residents	-0.83*	0.24
	(0.46)	(0.48)
Municipality with between 10,001 and 20,000 residents	-1.83***	-0.11
	(0.49)	(0.54)
Municipality with 10,000 or less residents	-2.01***	-1.09**
	(0.45)	(0.47)
Number of observations	5723	5296
Adjusted R-squared	0.40	0.52

Note: ***, **, * denote significance at 1, 5, and 10 percent level respectively; robust standard errors in parentheses

Table-3 (continued) Separate Estimations by Gender

Dependent Variable	(1) ISEI-M (OLS)	(2) ISEI-F (OLS)
Andalucía	*Reference Category*	
Aragón	0.652	0.87
	(0.79)	(0.84)
Asturias	1.28	0.31
	(0.96)	(0.95)
Baleares	1.73*	0.80
	(0.90)	(0.91)
Canarias	-0.66	-1.28
	(0.87)	(0.91)
Cantabria	0.25	-1.66*
	(0.88)	(1.00)
Castilla y León	0.23	1.04
	(0.74)	(0.83)
Castilla-La Mancha	1.27*	0.84
	(0.77)	(0.83)
Cataluña	1.82***	1.10*
	(0.66)	(0.67)
Valencia	0.48	-0.19
	(0.69)	(0.77)
Extremadura	-0.34	0.98
	(0.82)	(0.89)
Galicia	1.09	-1.04
	(0.77)	(0.82)
Madrid	1.49**	-0.09
	(0.64)	(0.71)
Murcia	-0.05	-1.51
	(0.80)	(0.93)
Navarra	-0.85	-0.52
	(0.93)	(0.94)
País Vasco	0.76	0.00
	(0.75)	(0.86)
La Rioja	-0.26	-0.39
	(0.89)	(0.95)
Ceuta	-1.62	-0.29
	(1.45)	(1.77)
Melilla	-0.39	-0.68
	(1.93)	(1.98)
Living with a partner	*Reference Category*	
Does not live with a partner	-1.74***	-0.32
	(0.43)	(0.40)
Number of observations	5723	5296
Adjusted R-squared	0.40	0.52

Note: ***, **, * denote significance at 1, 5, and 10 percent level respectively; robust standard errors in parentheses

Table-3 (continued) Separate Estimations by Gender

Dependent Variable	(1) ISEI-M (OLS)	(2) ISEI-F (OLS)
Number of co-living family members over 25 years old	-0.69***	-0.43**
	(0.21)	(0.21)
Number of co-living family members between 14 and 24 years old	-0.02	-0.28
	(0.22)	(0.22)
Number of co-living family members between 0 and 13 years old	-0.20	-0.04
	(0.21)	(0.23)
Number of observations	5723	5296
Adjusted R-squared	0.40	0.52

Note: ***, **, * denote significance at 1, 5, and 10 percent level respectively; robust standard errors in parentheses

6. Conclusions

This paper measures the returns to English proficiency by using ISEI-08 as the indicator. The standard OLS estimations show that workers with at least elementary English proficiency can get a higher ISEI-08 score than those without any knowledge of English (the base group), and such benefit grows gradually as their English skills improve. After correcting the auto-selection into the labor market bias, such effects do not experience an obvious change. Workers with advanced English proficiency in average gain 3.39 more ISEI-08 scores than the base group. This pattern remains the same for both male and female respondents without significant gender differences, which, however, should be taken with caution since no agreement has been arrived to on this topic yet. On the other hand, not surprisingly, individuals' education level constitutes the most influential control variable in this relationship. Even with just secondary education can increase the ISEI-08 score by at least 4 units relative to those with primary education or less. The effect of education is especially impressive when the respondents obtain at least a university degree, with an increase of over 20 units compared with primary education or less. Also, there is an obvious gender difference in terms of the returns to education level. Female workers benefit more than male workers from holding a higher education degree. At its most, female worker with at less university education share a 32.54-unit improvement in their ISEI-08 scores compared with those with primary education or less. Other factors included in the analysis in general do not present a strong effect, except that living in a municipality with less than 10,000 residents may incur at lease 1-unit decrease in ISEI-08 due to less labor opportunities.

This paper enriches the studies on the returns to English proficiency. More importantly, it targets at a non-English-speaking country, where more exact estimations

could be obtained as explained before. It also differs from similar studies by looking at the same problem from a new perspective, i.e., employing a new indicator (ISEI-08) whose potential has been recognized by several recent studies (Hamid & Nguyen, 2016; Pinilla-Portiño, 2018). The results are similar to what previous studies got using earnings as the outcome, which makes sense since earnings is an important construction of ISEI-08.

This paper acknowledges the benefits of learning English. In a way it explains the enthusiasm shown by institutions and individuals around the world for acquiring English skills. Apart from enhancing the general employability of people, English proficiency constitutes a fairly easy way to facilitate social mobility, which may give way to future subsidies on English learning in areas where a higher degree of social mobility is wanted. Also, this paper confirms the importance of education. Particularly, it discovers that female workers may benefit more from education than their male counterparts, based on which policymakers and individuals can evaluate better certain choices of educational investment.

Future researchers can try to validate the results from this paper in other countries, especially with cross-sectional data in order to analyze if the relationship varies across countries. Also, gender differences should be checked out more rigorously. Due to the common disagreements that exist in current studies, accompanied by the reasonings from Mora & Davila (1998), there lies a chance that gender differences in terms of the returns to English might be sectoral, i.e., in certain occupations, workers of either gender may benefit more from their English skills while in some others their English proficiency does not reward as much. Future studies may look into this hypothesis when a more detailed and suitable database is available.

References

Azam, M., Chin, A., Prakash, N. (2013). "The Returns to English-Language Skills in India", Economic Development and Cultural Change, 61(2), 335-367.

Bergman, M. M. & Joye, D. (2005). "Comparison social stratification schemata: CAMSIS, CSP-CH, Goldthorpe, ISCO-88, Treiman, and Wright", Cambridge Studies in Social Research, 10, 1-35.

Bleakley, H. & Chin, A. (2004). "Language skills and earnings: Evidence from childhood immigrants", The Review of Economics and Statistics, 86(2), 481-496.

Casale, D. & Posel, D. (2011). "English language proficiency and earnings in a developing country: the case of South Africa", The Journal of Socio-Economics, 40(4), 385-393.

Chiswick, B. R. & Miller, P. W. (1999). "Language skills and earnings among legalized aliens", Journal of Population Economics, 12(1), 63-89.

Connelly, R., Gayle, V., Lambert, P. S. (2016). "A Review of occupation-based social classifications for social survey research", Methodological Innovations, 19, 1-14.

Coxon, A. P. M. & Jones, C. L. (1978). The Images of Occupational Prestige. London: Macmillan.

Di Paolo, A. & Tansel, A (2015). "Returns to foreign language skills in a developing country: The Case of Turkey", The Journal of Development Studies, 51(4), 407-421.

Di Paolo, A. & Tansel, A. (2019). "English skills, labour market status and earnings of Turkish women", Empirica, 1-22.

Erikson, R. & Goldthorpe, J. H. (1992). The Constant Flux: A Study of Class Mobility in Industrial Societies. Oxford: Clarendon.

Erikson. R., Goldthorpe, J. H., Portocarero, L. (1979). "Intergenerational class mobility in three western societies: England, France, and Sweden", British Journal of Sociology, 30(4), 415-441.

Fabo, B., Beblavý, M., Lenaerts, K. (2017). "The importance of foreign language skills in the labour markets of Central and Eastern Europe: Assessment based on data from online job portals", Empirica, 44(3), 487-508.

Ganzeboom, H. B. G. (2010). “A new International Socio-Economic Index (ISEI) of occupational status for the International Standard Classification of Occupation 2008 (ISCO-08) constructed with data from the ISSP 2002-2007”, paper presented at the Annual Conference of International Social Survey Programme, Lisbon.

Ganzeboom, H. B. G., De Graaf, P. M., Treiman, D. J. (1992). “A standard International Socio-Economic Index of Occupational Status”, Social Science Research, 21(1), 1-56.

Ganzeboom, H. B. G. & Treiman, D. J. (2003). Advances in Cross-National Comparison. New York: Plenum.

Ginsburgh V. A. & Prieto-Rodriguez, J. (2011). “Returns to foreign languages of native workers in the EU”, Industrial & Labor Relations Review, 64(3), 599-617.

Ginsburgh, V. A. & Prieto-Rodriguez, J. (2013). “Is there a gender bias in the use of foreign languages in Europe?” Kyklos, 66(4), 552-566.

Hamid, M. O. & Nguyen, H. T. M. (2016). “Globalization, English language policy, and teacher agency: Focus on Asia”, The International Education Journal: Comparative Perspectives, 15(1), 26-44.

Heckman, J. J. (1976). “The common structure of statistical models of truncation, sample selection, and limited dependent variables and a simple estimator for such models”, Annals of Economic and Social Measurement, 5(4), 475-492.

International Labour Organization (2016). https://www.ilo.org/public/english/bureau/stat/isco/isco08/index.htm.

Mora, M. T. & Davila, A. (1998). “Gender, earnings and the English skill acquisition of Hispanic workers in the United States”, Economic Inquiry, 36(4), 631-644.

Nakao K. & Treas, J. (1994). “Updating occupational prestige and socio-economic scores: How the new measures measure up”, Sociological Methodology, 24, 1-72.

Pinilla-Portiño, N. (2018). “The Influence of learners’ socio-economic status on learning English as a foreign language”, The Journal of Asia TEFL, 15(2), 550-558.

Shields, M. A. & Price, S. W. (2002). “The English language fluency and occupational success of ethnic minority immigrant men living in English metropolitan areas”, Journal of Population Economics, 15(1), 137-160.

Wang, H., Smyth, R., Cheng, Z. (2017). “The economic returns to proficiency in English in China”. China Economic Review, 43(C), 91-104.

Williams, D. (2011). “Multiple language usage and earnings in Western Europe”, International Journal of Manpower, 32(4), 372-393.

Appendix: Descriptive Statistics of Variables

A. Summary of continuous variables for working respondents

Count	Mean	S.D.	Min.	Max.
ISEI-08				
11,019	41.70	15.41	10	69
Age				
11,019	44.41	10.54	18	64
Number of co-living family members over 25 years old				
11,019	2.16	0.76	0	10
Number of co-living family members of 14 - 24 years old				
11,019	0.45	0.72	0	10
Number of co-living family members of 0 - 13 years old				
11,019	0.56	0.83	0	10

B. Summary of continuous variables for non-working respondents

Count	Mean	S.D.	Min.	Max.
ISEI-08				
0				
Age				
4,599	46.08	12.45	18	64
Number of co-living family members over 25 years old				
4,599	2.29	0.85	0	9
Number of co-living family members of 14 - 24 years old				
4,599	0.49	0.74	0	4
Number of co-living family members of 0 - 13 years old				
4,599	0.41	0.77	0	6

C. Summary of dummy variables for working respondents

Dummies	Frequency	Percentage	Cumulative
Level of English proficiency			
No English skills	6479	58.8	58.8
Basic skills	779	7.07	65.87
Elementary skills	1092	9.91	75.78
Intermediate skills	1741	15.8	91.58
Advanced skills	928	8.42	100
Highest education obtained by parents			
Primary education or less	5551	50.38	50.38
Secondary education	2484	22.54	72.92
Tertiary education	1356	12.31	85.23
University or higher	1628	14.77	100
Total	11019	100	100

C. (continued) Summary of dummy variables for working respondents

Dummies	Frequency	Percentage	Cumulative
Highest education obtained by respondent			
Primary education or less	186	1.69	1.69
Secondary education	3188	28.93	30.62
Tertiary education	4237	38.45	69.07
University or higher	3408	30.93	100
Size of the municipality			
Capital of province or with more than 100.000 residents	5409	49.09	49.09
Municipality with between 50,001 and 100,000 residents	884	8.02	57.11
Municipality with between 20,001 and 50,000 residents	1651	14.98	72.09
Municipality with between 10,001 and 20,000 residents	1229	11.15	83.25
Municipality with 10,000 or less residents	1846	16.75	100
Autonomous communities			
Andalucía	1324	12.02	12.02
Aragón	592	5.37	17.39
Asturias	402	3.65	21.04
Baleares	376	3.41	24.45
Canarias	441	4	28.45
Cantabria	362	3.29	31.74
Castilla y León	692	6.28	38.02
Castilla-La Mancha	584	5.3	43.32
Cataluña	1114	10.11	53.43
Valencia	848	7.7	61.12
Extremadura	483	4.38	65.51
Galicia	576	5.23	70.73
Madrid	1146	10.4	81.13
Murcia	475	4.31	85.44
Navarra	409	3.71	89.16
País Vasco	609	5.53	94.68
La Rioja	431	3.91	98.59
Ceuta	81	0.74	99.33
Melilla	74	0.67	100
Whether living with a partner			
Yes	7935	72.01	72.01
No	3084	27.99	100
Total	11019	100	100

D. Summary of dummy variables for non-working respondents

Dummies	Frequency	Percentage	Cumulative
Level of English proficiency			
No English skills	3370	73.28	73.28
Basic skills	296	6.44	79.71
Elementary skills	293	6.37	86.08
Intermediate skills	412	8.96	95.04
Advanced skills	228	4.96	100
Highest education obtained by parents			
Primary education or less	2940	63.93	63.93
Secondary education	869	18.9	82.82
Tertiary education	409	8.89	91.72
University or higher	381	8.28	100
Highest education obtained by respondent			
Primary education or less	276	6	6
Secondary education	2411	52.42	58.43
Tertiary education	1358	29.53	87.95
University or higher	554	12.05	100
Size of the municipality			
Capital of province or with more than 100.000 residents	2121	46.12	46.12
Municipality with between 50,001 and 100,000 residents	381	8.28	54.4
Municipality with between 20,001 and 50,000 residents	706	15.35	69.75
Municipality with between 10,001 and 20,000 residents	514	11.18	80.93
Municipality with 10,000 or less residents	877	19.07	100
Autonomous communities			
Andalucía	853	18.55	18.55
Aragón	169	3.67	22.22
Asturias	187	4.07	26.29
Baleares	114	2.48	28.77
Canarias	297	6.46	35.23
Cantabria	176	3.83	39.05
Castilla y León	249	5.41	44.47
Castilla-La Mancha	273	5.94	50.4
Cataluña	330	7.18	57.58
Valencia	377	8.2	65.78
Extremadura	295	6.41	72.19
Galicia	221	4.81	76.99
Madrid	316	6.87	83.87
Murcia	207	4.5	88.37
Navarra	103	2.24	90.61
Total	4599	100	100

D. (continued) Summary of dummy variables for non-working respondents

Dummies	Frequency	Percentage	Cumulative
País Vasco	170	3.7	94.3
La Rioja	129	2.8	97.11
Ceuta	72	1.57	98.67
Melilla	61	1.33	100
Whether living with a partner			
Yes	3003	65.3	65.3
No	1596	34.7	100
Total	4599	100	100