English Skills and Wages in a Non English Speaking Country: Findings from Online Advertisements in Mexico

Miguel Delgado Helleseter*
University of California, Santa Barbara

March 22, 2013

*Address: 2021 North Hall, Department of Economics, University of California, Santa Barbara, California, 93106-9210, USA, telephone: +1-805-765-1623, e-mail: mdh@econ.ucsb.edu

Abstract

In spite of the generally accepted status of English as a *lingua franca*, the labor market returns to English for its role as an international language are understudied. In this paper I use advertisements from Computrabajo.com.mx to estimate the returns to English in Mexico. I find that the wage premium for English speakers is approximately 28 percent for the sample as a whole. The evidence suggests that English may provide a path to upward economic mobility in Mexico, and that this holds true for persons of all skill levels.

Keywords: language, human capital, skill, wage

JEL Classification Numbers: J24, J31, O15

1 Introduction

The labor market value of English has been studied in numerous articles, yet it is usually studied in countries where English is the primary language of business and government. As a result, much of this literature is closely related to immigration and the value of English skills for immigrants to English-speaking countries. In that context, English can be viewed as a host-country-specific skill and its measured value can be interpreted as the admission ticket to the national economy.

In today's global economy, English has an additional and important source of value because of its role as an international language. However, in spite of its generally accepted status as a *lingua franca*, there is surprisingly little evidence on the value of English as an international language. This is in part due to lack of good data on English ability in countries where English is not an official language. Language-earnings studies are generally conducted using household surveys, and, in countries where English is not widely used for business and/or government, household survey data on English proficiency is scarce at best.

This paper fills the gap in the literature by collecting a new type of labor market data that, with some work, is readily accessible in almost any country: internet job board data. Specifically, I use Mincer-type regressions to estimate the labor market value of speaking English using a large and unique dataset of advertisements from Computrabajo.com.mx, a Mexican internet job board. Using job board data has the advantage of offering a large sample of data with information that is rarely available in surveys. This is particularly true for information on English skills in countries like Mexico, where English is not widely used for business and/or government.

While job board data applies to a select part of the population rather than to the entire labor market of a country, and even though it lacks some useful controls typically available in social surveys (such as workers' family background measures, for example), job board data

does offer three particular advantages over survey data for the study of language skills and wages.

First, job board data is arguably less vulnerable to measurement error than survey data as the latter suffers from at least two problems: individuals don't know their ability or use different scales to determine whether they are proficient or not, leading to classical measurement error. Also, individuals tend to exaggerate their ability, which leads to them concentrating in the higher categories. Both of these problems lead to attenuation bias. Wages advertised on job boards, on the other hand, should apply only to workers whose English language ability is good enough to land them a job offer.

A second advantage is that, unlike most social and labor market survey data, I can control for employer, and employer*occupation fixed effects. This is important because estimates of the returns to English using survey data may be upward biased if those reporting to be English-proficient disproportionately work in firms that pay high wages. With firm fixed effects I can compare offered wages in advertisements posted by the same firm, for jobs in the same occupation, when the firm requests an English speaker and when it does not. Thus, controlling for employer fixed effects corrects for spurious effects that may arise from correlations between English and employer type. In addition, since firms would most likely not pay a premium for a skill they do not consider necessary, my approach may provide a more direct measure of the value of English *actually used* on the job than do most studies using social survey data.[1]

Lastly, I have good controls for unobserved ability. In this paper I exploit a unique feature of the job board data to reduce unobserved ability bias by controlling for a number of variables associated with cognitive and non-cognitive skills that are not available in survey data. With the exception of a small number of studies of new-language-learners that use individual fixed effects to control for time-invariant unobserved ability, existing studies of language skills and earnings have few controls for cognitive and non-cognitive skills.[2] Furthermore, studies using person fixed effects apply only to a highly selected sample of persons who choose to learn English as a second language relatively late in life.

The English premium I estimate (within firm-occupation-state cells) in my fully specified model is large, at approximately 28 percent for the sample as a whole. That is, observationally identical jobs with a requirement for English skills or bilingualism advertise a wage

[1] Two exceptions using data from the Canadian census on primary and secondary languages used at work are Christofides and Swidinsky (2010) and Grenier and Nadeau (2011).

[2] Lang and Siniver (2009) effectively use person fixed effects by using first differences. However, as the authors point out, there are concerns with measurement error in self-reported proficiency levels, and even after differencing errors could remain due to, for example, lack of confidence in one's language skills at the time of the first survey and over-confidence at the time of the second survey.

that is 28 percent higher than similar jobs in the same firm and occupation without this requirement. This effect is roughly equivalent to the estimated wage difference associated with having a college degree compared to having an incomplete college education.

I also explore whether a complementarity exists between language acquisition and other skills, a common question in the language-earnings literature. I find no evidence of complementarity between language and education or experience. Jobs requiring the lowest levels of education show statistically higher returns to English language skills than jobs with higher education requirements. Also, the English wage premia for jobs requiring more years of experience or requiring supervision and management of personnel are not statistically different from the English wage premia for jobs without these requirements.

2 Background and Related Literature

2.1 Background

This study focuses on an area of the language-earnings literature that has been studied relatively little, namely that of the returns to second-language skills in general, but more specifically the returns to English, a World language, in a non-English speaking country. It is unquestioned that English serves as an international language and that its role as such is growing in much of the world. Mexico is a prime example.

According to statistics from Mexico's office of the Secretary of the Economy (2012*a*), Mexico's total exports have been growing at an annual average rate of 11 percent since 1993, and throughout that period an average of 84 percent of these exports have been to the United States. As of 2011, Mexico ranked among the top three trading partners of the United States in total volume of both imports and exports of goods (United States Census Bureau, 2012). Although large trading volumes alone may not be sufficient to make English a particularly valuable skill in Mexico, there is evidence that suggests that the demand for English skills in Mexico has been growing in recent years. While total Foreign Direct Investment (FDI) into Mexico has remained fairly flat over the past decade, as has FDI into some of the traditional investment sectors like manufacturing, FDI in sectors that are more likely to require English speakers, such as professional services or lodging, have seen relatively strong growth over the past several years (*Secretaría de Economía, México*, 2012*b*). Furthermore, growth in employment in these sectors over the past seven years has far exceeded the country's overall growth in employment(*Instituto Nacional de Estadística y Geografía*, 2012).

Recognizing that English is becoming an increasingly important tool for the Mexican workforce, the Mexican government began a program in 2009 to introduce English education

to all levels of basic education (*Secretaría de Educación Pública, México*, 2007). This plan, however, is just finishing its pilot stage and is therefore not widespread throughout Mexico yet. Furthermore, the labor market outcomes of this plan will not begin to crystalize for perhaps another decade. Currently there is little information on the use of English in Mexico, but a recent survey (*Consulta Mitofsky*, 2007) suggests that less than 5 percent of Mexicans speak English well. This is important because ultimately any premium for being bilingual will depend on the interaction of supply of, and demand for, English skills.

In spite of the government's recognition of the importance of English as a skill for the Mexican labor force, there is, to my knowledge, no research on the magnitude or measured value to the returns to English in Mexico. This is true not only for Mexico but also for many countries where English has become increasingly important as an international language. This research gap is partly due to lack of good data on English ability. In this study I fill the gap in both the literature and the data by exploring the relationship between second language proficiency and wages in Mexico using a novel data set from an internet job board.

2.2 Second-language proficiency

Immigration and language

Given the relatively small number of papers on the topic of foreign-language proficiency and earnings, it is perhaps useful to review some of the literature in the related field of host-nation language skills of immigrants and earnings, which was the focus of much of the language-earnings literature in its initial stages.[3] In the 1980s, McManus, Gould and Welch (1983) and Grenier (1984) found a connection between language skills and wages for Hispanic men in the United States, and Tainer (1988) extended this research to include all foreign-born men.

In the 1990s, McManus (1990) studied the effects of Spanish-speaking enclaves on earnings, while Dustmann (1994) used German data and expanded the analysis to include females and to include variables in the analysis such as degree of literacy. Using census data from Israel, Chiswick (1998) examined the determinants of, and the returns to, Hebrew fluency for male immigrants. Chiswick's paper is also one of the first to examine returns to English in a non-English speaking country, where he found that Hebrew speakers who speak English as a second language have the highest earnings of all people in his analysis.

Recent work by Borjas and Friedberg (2009) examines long term trends in immigrants' earnings and potential causes for changes in the trends. Other recent work attempts to

[3]The literature on immigrant assimilation is extensive. For surveys and more extensive review than that provided here see LaLonde and Topel (1997), Borjas (1999) and Bleakley and Chin (2004)

explain some of the mechanisms by which immigrants' wages are affected by English proficiency. Chiswick and Miller (2010) in one of the more current papers in this area find that foreign-born workers' compensation is positively related to required English proficiency of their jobs, and they find that correctly matching English proficiency level with the proficiency requirement of the job is also rewarded.

Value of foreign language

Some researchers have investigated the link between second-language proficiency and earnings in countries where that second language is either an official language or a de facto language used extensively in business and government. This includes research on French language skills in Morocco by Angrist and Lavy (1997), on English and French in Canada by Christofides and Swidinsky (2010), as well as English in South Africa (Casale and Posel, 2011*a*; Casale and Posel, 2011*b*). The case of Mexico is different in that English is not used extensively in government and business, yet is potentially important because of the geographic proximity to the United States and the trade relationship between the two countries.

In this regard, this current research is more in line with the work by Saiz and Zoido (2005), Munshi and Rosenzweig (2006), Lang and Siniver (2009), Azam, Chin and Prakash (2010), and Ginsburgh and Prieto Rodriguez (2011), although some differences in scope and approach exist. Saiz and Zoido (2005) explore the premium in the United States for speaking a second language besides English. Part of the motivation for my research is the generally accepted view of English as a *lingua franca*, so to that extent the scope of these analyses differ.

Munshi and Rosenzweig (2006) estimate returns to English in a non-English speaking country (India), yet as pointed out by Azam et al. (2010), these are returns to English education rather than to English skills as they explore sorting into occupations based on choice of schooling (where a key difference is English education). Lang and Siniver (2009) examine the return to English for natives and Russian immigrants in Israel and use individual fixed effects to control for unobservables. Azam et al. (2010) estimate the returns to English skills in India using a newly released India Human Development Survey and find significant returns to English and evidence supporting language-skill complementarity.

Lastly, Ginsburgh and Prieto Rodriguez (2011) study the return to non-native language skills in Europe, i.e. the returns to speaking non-native languages, one of which is English, for a number of European countries. The authors find significant positive returns to non-native language skills and find that while English commands higher returns in northern Europe, other languages may be more highly rewarded in southern Europe (such as French in Spain).

Most of the research in this area relies on surveys and consequently on self-reported language proficiency levels (with the exception of Ginsburgh and Prieto Rodriguez who use reports on language used at work). My paper adds to the literature by providing a different approach that does not rely on self-reported data.

Estimation challenges

Over the last two decades, much attention has focused on possible solutions to two problems in the language-earnings literature, specifically potential unobserved heterogeneity (ability bias), and potential measurement error in the language proficiency measures, which were almost entirely from self-reported data. Several authors have addressed the unobserved ability bias problem using instrumental variables, including work by Chiswick and Miller (1995), Dustmann and Van Soest (2002), and Bleakley and Chin (2004).

Using census data from Australia to estimate returns to English for immigrants, Chiswick and Miller (1995) use three instruments for English fluency, whether the immigrant was married overseas, the number and ages of children, and a concentration measure of non-English language in the immigrant's region of residence. Dustmann and Van Soest (2002) examine the relationship between immigrant host-country language proficiency and earnings using a ten-year panel in Germany, instrumenting for language proficiency using parental education variables. In both cases, while it is reasonable to argue that the chosen instruments should not enter the earnings equation directly, it is still plausible that these variables affect wages through channels other than language fluency.

Bleakley and Chin (2004) convincingly instrument for English proficiency of immigrants to the United States using age at arrival interacted with a dummy variable for non-English speaking country of origin. The age of arrival is used under the argument that younger children learn languages more easily. Recognizing that age of arrival may operate through channels other than language, the authors use immigrants from English-speaking countries as a control for non-language related mechanisms in which age of arrival may affect earnings.

The problem of measurement error from self-reported language proficiency has been documented — see Dustmann and Van Soest (2002) and Bleakley and Chin (2004), and some offered solutions include using interviewer-assessed proficiency levels (Hamilton, Goldsmith and Darity Jr., 2008) or a combination of self-reported and interviewer-assessed proficiency levels (Redstone Akrehs and Frank, 2011). According to Hamilton et al. (2008), the difference in language wage premia estimates resulting from using interviewer-reported rather than self-reported measures of language proficiency ranges from less than one percentage point and up to ten percentage points, depending on gender and the proficiency criteria used.

Using wages and self-reported English proficiency levels from the New Immigrant Survey, Redstone Akrehs and Frank (2011) estimate and compare language wage premia between all respondents and a sub-sample of these respondents whose self-reported proficiency is consistent with an interviewer-reported proficiency evaluation. The authors find that for the sub-sample, the estimated premium is larger than that for the whole sample by roughly 13 percentage points.

In this paper I address issues of ability bias and measurement error by exploiting unique features of the job board data. First, I am able to include firm fixed effects, which controls for the possibility that high-wage firms disproportionately request English speakers. In addition, I have a number of control variables associated with cognitive and non-cognitive skills. In regards to measurement error, I argue that the specific requirement for an English speaker as well as any proficiency level requested by a firm should reflect their actual need, and therefore might be more precise than self-reported bilingualism or proficiency levels from survey data.

2.3 Internet job board studies

This study uses data from an online job board, a relatively new and promising data source for researchers interested in a variety of labor market questions. Different aspects of the impact of the internet on labor markets have been examined in previous studies. Kuhn and Skuterud (2000) compared traditional job search methods to then-relatively-new internet job search methods, and Autor (2001) discussed three ways in which the labor market could be affected by the internet. Within this literature, however, the use of internet job boards to examine labor markets is fairly new and much less explored, which is not surprising given that it is a relatively new source compared to other job advertisement and search methods such as newspapers or employment agencies. In addition, as Stevenson (2008) points out, job boards are just one of the many ways in which the internet is used for job search purposes.

While this current study uses advertised wages from job board data to estimate wage (Mincer-type) equations, the few studies I know of that use online job board data have focused on different questions. Brencic and Norris (2009; 2010*a*; 2010*b*) analyzed several aspects of employer search for workers in a series of papers using data from job site Monster.com (among other sources). Brencic (2011) also examined the decision by firms of whether to post a wage (or range of wages) in their advertisements. Kuhn and Shen (2013; 2012) used data from an online job board in China to examine gender discrimination and the preferences of employers for gender, age, height, and beauty, as well as firm hiring practices based on qualifications (and over qualifications) of potential employees. Using data from Career-

builder.com, Marinescu (2012) studies online job search and unemployment insurance, and Marinescu and Wolthoff (2012) study whether higher posted wages attract more and better (higher-skilled) candidates.

3 Data

This study uses data from job advertisements posted for employment in Mexico on Computrabajo.com, an online job board with job postings in Latin American countries and Spain. The data used in this study consists of the universe of unique advertisements posted on the company's Mexican site (Computrabajo.com.mx) during the approximately 18-month period from January 25, 2011 through August 25, 2012.[4] Unique advertisements include advertisements that appeared once during this time, as well as identical advertisements appearing on multiple dates (in which case only the first appearance is included in the sample). Additional details about how advertisements were collected and filtered are provided in Appendix C.

Since my approach is to estimate wage regressions, the sample I analyze naturally contains only those observations that clearly state a wage or salary. Also, since it is unquestioned that education level is an important determinant of labor market outcomes, I restrict my sample to include only observations in which the required level of education is explicit. Lastly, many advertisements on Computrabajo.com are posted by staffing firms, and firm fixed-effects in this analysis are estimated using a firm variable that reflects the firm posting the advertisement, which, in the case of staffing firms, differs from the actual firm that will potentially provide employment. To provide a cleaner interpretation of all controls, the sample I analyze includes only firms whose primary function is not that of providing staffing services to other firms.

Advertised salaries were determined as the average of the minimum and maximum advertised salary for each advertisement, where the minimum and the maximum were the same if a salary rather than a range was advertised. Advertisements with monthly salaries equivalent to 1,700 Mexican pesos or less were dropped from the analysis since these would be essentially below minimum wage (see Appendix C). Advertisements with salaries above 100,000 Mexican pesos per month were also dropped, as were advertisements where the salary range (maximum-minimum) was more than double the minimum.[5]

A total of 202,748 unique advertisements were posted during the 18-month data collec-

[4]For a period of approximately one and a half months from May 14 to June 24, 2012 data were not collected properly and these observations were therefore dropped from the analysis

[5]Many advertisements indicate that the advertised salary either includes or does not include benefit payments, bonuses, commissions, and other extras. For the analysis presented here, ranges were created based on the advertised amounts, ignoring this additional language.

tion period. Of these, 44.3 percent (89,874 observations) have an explicit education requirement and 24.2 percent (49,100 observations) state a salary that falls within the parameters described above. Only 24,721 observations (12.2 percent of the whole sample of unique advertisements) state both a salary and a required education level. Of these, approximately 56.2 percent are posted by firms whose primary business is to provide staffing services to other firms, thus yielding a total of 10,840 observations that satisfy my sampling criteria. See Appendix C for a discussion of the impact of imposing these sample restrictions.

As shown in Table 1, slightly less than ten percent of advertisements are for jobs requiring English-speaking skills. A majority of advertisements (55.4 percent) are for jobs requiring at most a high school diploma. Jobs requiring some postsecondary education account for 15.0 percent of the total and the additional 29.6 percent consists of jobs requiring a university degree.[6] Almost 85 percent of all advertisements have an explicit age requirement (30.3 years of age on average) and approximately one half has explicit minimum experience requirements (1.1 years of minimum experience on average).[7] The average advertised salary is 7,180 Mexican pesos per month (approximately 568.40 U.S. dollars as of the time of this writing), well above average salaries in Mexico.[8]

All advertisements on Computrabajo.com are categorized into a set of 14 broad occupations, shown in Table 2. Interestingly, despite how broadly defined these occupations are, there is enough overlap between them that many jobs are classified under more than one of these occupations, and these observations are therefore duplicates. The data in Table 2, which excludes all such duplicates, shows that jobs in sales as well as health care-related jobs and those classified as other have the lowest share of advertisements with requests for English speakers, all below 5 percent. On the other end are jobs in information technology, where 25.4 percent of advertisements request English-speaking candidates. Other occupations with relatively high requests for English speakers include engineering or technical jobs and jobs in education or academia. Advertisements requesting English-speakers appear more often for jobs requiring higher education levels (see Table 3). While less than 2 percent of advertisements in the lowest education levels request English speakers, about 8.2 percent of advertisements requiring some college or a technical degree do request English speakers, and

[6]Education requirements were collected at a finer level but were aggregated to the levels discussed here for analytical purposes because of otherwise small education cells. High school or less includes observations requiring a high school diploma, as well as observations with education requirements lower than high school completion.

[7]Age requirements were often requested using a range rather than a unique value. For all observations with age requirements a range was generated and the average of the range was used. Age was not calculated for observations with only a minimum or maximum age request.

[8]Based on reports to the Mexican social security administration, the average salary in Mexico is close to 5,000 Mexican pesos per month, or about 396 U.S. dollars per month (*Instituto Nacional de Estadística y Geografía*, 2012)

this increases to 22.1 percent for advertisements requiring a bachelor's degrees or higher.

This sample of advertisements from Computrabajo.com is not representative of all jobs in Mexico. To illustrate this point, Table 4 presents summaries of salaries, age, and education for the advertisements that include this information and for official statistics.[9] Advertised salaries on Computrabajo advertisements are about 44 percent higher than the estimated average wage in Mexico based on employer reports to the Mexican social security administration. At the same time, advertisements that request candidates in a specific age range are on average requesting candidates who are younger than the average employed worker. More importantly, perhaps, is that advertisements on Computrabajo.com have education requirements that are on average well above average education levels for the population. While almost two thirds of the population have completed secondary or less, advertisements requesting candidates with such education levels appear only about 20 percent of the time, while advertisements requesting education of high school and above account for 78.4 percent of all advertisements.

In addition, advertisements on Computrabajo are heavily concentrated in the *Distrito Federal* area, accounting for 47.4 percent of all advertisements (excluding work abroad and work classified as "whole country"), in contrast with actual population which is much more spread over the country. Although some of these differences may be attributable to the *Distrito Federal* being more of an employment center than a housing area, even when we add population in the state of Mexico (*Estado de Mexico*) to the *Distrito Federal*, the share of the population these represent is slightly above one fifth, whereas the share of advertisements for *Distrito Federal* and *Estado de Mexico* combined account for close to two thirds of all advertisements. Furthermore, some of the differences between requested education levels and the population's actual education levels persist even when looking within *Distrito Federal* and *Estado de Mexico* (in fact this holds for all states). The fact that within all states explicit education requirements are well above the average education levels suggests that, as expected, advertisements on Computrabajo are on average for jobs requiring higher skills and education levels than what is found on average in the general population.

There are some additional sampling issues worth pointing out. A company with no positions to fill currently or in the near future would be very unlikely to post a vacancy, whereas firms with many vacancies would be likely to advertise positions. Therefore it is likely that advertisements on Computrabajo (or in any job board for that matter) contain a larger share of vacancies from companies with high employee turnover or companies that

[9]Official statistics are salaries based on reports to the Mexican social security administrator as well as population and employment data from the 2012 National Employment and Occupation Data survey(*Instituto Nacional de Estadística y Geografía*, 2012).

are growing and adding jobs than from companies with lower turnover or in non-expanding industries. Similarly, the decision to post specifically on Computrabajo as opposed to other job advertisement outlets is also endogenous.

A similar within-sample issue is that posting a wage (and therefore being included in the sample analyzed) is determined endogenously. Firms posting wages may be systematically different from firms not posting wages. This could include differences in actual salaries paid between wage-posters and non-wage posters. According to Brencic (2011), a consistent empirical finding from her research using three different job vacancy sources (one on-line, one in a career center, and one in an employment agency) is that the likelihood of posting a wage decreases with the skill requirement of the job.

Whether endogeneity affects my estimates of the returns to English depends both on whether jobs posted on Computrabajo are positively or negatively selected (on wage) relative to all jobs (in which case there is attenuation bias), and on whether the English premium is higher than average in Computrabajo jobs. If jobs advertised on Computrabajo have a higher premium, I will overestimate the English premium, otherwise, I will underestimate it.

4 Estimation results

To measure the value of speaking English in Mexico I use data from Computrabajo advertisements to estimate Mincer-type wage regressions (Mincer, 1974) of the form

$$Ln\ wage_i = \alpha + \beta_0 English_i + \beta_1 Education_i + \beta_2 Experience_i + \gamma X_i + u_i$$

where i indexes each advertisement and X is a vector of observable characteristics such as age and gender and $English$ is an indicator (dummy) variable for whether the advertisement requests a candidate who speaks English (or is bilingual).

A noteworthy feature of the vector X is that it includes observable characteristics that are not available in the survey data typically used in studies of earnings and language skills, several of which can be associated with unobserved factors that may bias the results in studies using survey data where information about these particular characteristics or traits are not available. The argument is that salaries and one or more of the covariates may be correlated with unobserved ability, where ability may lead to both higher salaries as well as to being more likely to learn a second language, which results in an upward bias of this coefficient since it is in fact picking up not only the English effect but also the effect of ability.

Data from advertisements provide an advantage over other data used in the language-

earnings literature in that it allows me to construct indicators of cognitive and non-cognitive skills needed in the job based on advertised salaries and explicit requests for these skills in the advertisements. Key words and particular ways of requesting certain skills were identified from a random sample of approximately 1,000 advertisements, and 36 specific non-cognitive skills and four cognitive skills were identified. Henceforth I will refer to these as cognitive and non-cognitive skill demands. Additional information about these skill demands can be found in Appendix C. A caveat for interpreting results from these data is that the absence of a mention of a particular skill in one advertisement does not imply that a particular skill is not important for the company posting that advertisement. The presence of key words associated with a particular skill should, nonetheless, indicate that such a skill is desired intensely enough by the company since the inclusion of the request in the advertisement should disqualify potential applicants who do not have (or do not believe they have) this particular trait.[10]

4.1 Main effects

Table 5 presents the results of a regression of a wage variable (*ln_salary*) on an indicator variable (*English*) equal to 1 if the advertisement included words such as English, *Inglés*, bilingual, or *bilingüe*, and equal to 0 otherwise. Based on 10,840 observations analyzed, the first column shows that prior to controlling for any other variables, a positive and significant difference exists between advertisements requesting candidates who speak English and those that do not. The coefficient on English (0.532) would represent a 70 percent increase in monthly salary from a predicted 5,873 Mexican pesos (465 U.S. dollars) per month for non-English speakers to 9,993 Mexican pesos (791 U.S. dollars) per month for English speakers. As should be expected, much of this effect may be working through channels such as education and occupation.

Once I include the standard wage regression variables (quadratics in age and experience, and indicators for education level and gender) the coefficient on English drops and the estimated premium is approximately 31 percent. The coefficients on age and experience show the expected signs, as do the education variables (where the omitted category are jobs with education requirements below high school, so we should expect all education coefficients to be positive and increase with education level).[11]

[10] While person fixed effects may provide better controls for time-invariant unobserved ability than these cognitive and non-cognitive skill demands, person fixed effects also have their limitations in that their reduction of biases in the coefficient of interest (English) only applies to changers in the data. That is, controls using person fixed effects are only relevant for observations switching from non-English speaking to English speaking.

[11] While the inclusion of age and experience in column 2 increase the English coefficient marginally, the

In the third column I include a dummy variable for whether the job requires supervision of personnel. I also include a control for whether the job is posted as full-time, dummy variables for requests for cognitive and non-cognitive skills, and state fixed effects. The addition of the supervision variables, the skill demand variables, the full-time job indicator, and the state fixed effects have a very small effect on the English coefficient. The inclusion of occupation controls in the fourth column also has a small impact on the English coefficient, as do the inclusion of company fixed effects and interactions with occupation and state (fifth through seventh columns). That is, after including the most basic controls (standard wage regressors), there is relatively little attenuation on the English coefficient.

This is true even in the last column, which has firm*occupation*state fixed effects. In other words, when comparing two advertisements by the same firm, for the same occupation, in the same state, one with a requirement for English speakers and one without this requirement, and controlling for age, gender, experience, cognitive and non-cognitive skill demands, and requirements to manage and/or supervise personnel, I still find a 28 percent English premium (English coefficient of 0.248). This premium is large, particularly considering that English is not an official language nor used extensively in government and business in Mexico. The estimated English premium is roughly equivalent to the estimated wage difference associated with having a college degree compared to having an incomplete college education.

4.2 Language-skill complementarity

In the language-earnings literature it is often found that a complementarity exists between skills and language acquisition. That is, the jobs where language skills are more valuable are higher-skill jobs. To explore this using the Computrabajo advertisements, I estimated models using interaction terms between the English variable and three different variables to proxy for skill: education, experience, and whether the job requires supervision or management of personnel.

The results shown in Table 6 provide no evidence of complementarity between language and skill. The interaction of English and education levels (second column) indicates that, if anything, the English premium is larger for jobs with a requirement for a high school diploma or less than for jobs with higher education requirements. Figure 1 displays these results graphically.

As a different proxy for skill, in the third column I interact English with a dummy variable for whether the job requires supervising or managing people. The coefficient on the interaction term is not statistically different from zero, while the coefficients on both

large drop in the coefficient is due to inclusion of the education variables.

the English and the require supervision variables are only slightly affected (compared to the main specification in the first column).

In the fourth column, I interact English with the quadratic on the minimum experience requirement. Experience is valuable as it improves skills, so experience should be a reasonable proxy for skill. The interaction terms are small and not significant, while the coefficient on English is only marginally larger than that in the main specification (first column). The results in the last column include interactions of English with all three of the proxies for skill. The patterns, magnitudes, and significance levels of the coefficients on all interactions are similar to that found in the columns with the individual interactions.

The evidence from Computrabajo advertisements suggest that, through the ability to speak English, employees in jobs with low requirements for other skills (particularly as measured by education), may "catch up" economically to employees in jobs with otherwise higher skill requirements.

4.3 Propensity score matching

I have attempted to reduce potential omitted variable bias by including a number of reasonable control variables in the regression analysis. An alternative approach to address this concern is to use propensity score matching. For the requirement for English speakers, one can estimate the propensity score as the probability that an advertisement requests English speakers as a function of certain (observed) characteristics. Then, each observation with a request for English speakers can be matched to an observation that has a similar propensity score but that does not have a request for English speakers, and the average difference in advertised salary for all matched pairs can be calculated.

Since only matched observations are compared, there is not much one can say about the unmatched observations, and the interpretation of the estimated salary difference is not the same as in the regression analysis. Matched advertisements without a requirement for an English speaker can be thought of being advertisements that are as likely as their matched pair to have a request for English speakers but that somehow did not have that request. Therefore, the English premium estimated this way is interpreted as the English premium for jobs that could have a requirement for English speakers.

While it shares the limitations of ordinary least square regression in that only observable variables can be included in the estimation, compared to regression, propensity score matching offers the advantage of reducing the reliance on linearity assumptions and reducing the concerns with over-fitting the model. Furthermore, propensity score matching may be used as a complement to regression analysis, where, as described by Morgan and Winship (2007),

matching may serve as a preprocessor of the data for further analysis using regression.

As Figure 2 shows, there is not much overlap of salaries between advertisements requesting English and those that do not. Using the English request binary variable as the dependent variable I estimated a logit model using the following covariates: age, experience, gender, indicators for whether age and experience are missing, indicators for education level, occupation indicators, an indicator for whether the job requires supervising or managing employees, and state indicators.

These covariates can reasonably be expected to affect both the need for English speakers as well as the advertised wage. At the same time, the requirements for age, gender, experience, education, and supervisorial skills, as well as the actual occupation and state of the job, can reasonably be expected to be unaffected by the need for English speakers. That is, it is difficult to imagine, for example, that after deciding whether a job requires an English speaker or not, the firm decides under which occupation to list the advertisement or in which state the job will be.

Using the results of the logit model I predicted the propensity scores for requesting English. While there is not much overlap of these scores between advertisements requesting English and those that do not for the sample as a whole (see Figure 3a), there is a lot of overlap of the propensity scores once we include only matched advertisements (see Figure 3b).[12] Using the matched advertisements I estimate the average salary difference to be a statistically significant 0.288, slightly larger than the result from the main regression specification.

The second column in Table 7 shows the results of the main specification using only observations matched on the propensity score. The coefficient on English (0.235) is only slightly smaller than in the main specification (first column). I conclude that, perhaps due to the rich set of covariates available for my main regression analysis, the concerns motivating the propensity score matching do not seriously bias those estimates. Finally, the fourth column shows the results of the specification that interacts English with the skill proxies using the propensity-score-matched sample. The English premium for the jobs with the lowest skill requirements is attenuated compared to the same specification using the whole sample (third column), yet the patterns for the covariates are similar to those for the specification using the whole sample.

[12]Each advertisement requesting English speakers was matched to the non-English-requesting advertisement with the closest propensity score. A total of four English-requesting advertisements were unmatched due to lack of a proper match. Also, some non-English-requesting advertisements served as a match for more than one English-requesting advertisement as the match was performed with replacement.

5 Conclusion

The returns to English for its role as an international language have been understudied. A number of studies examine the returns to English in countries where English is either an official or a de facto language used extensively in government and/or business, but such studies answer a similar question to that of the value of English skills for immigrants to English-speaking countries. In addition, while a few studies do explore the link between English skills and labor market outcomes in countries where English is not an official language, none of these countries share the specific characteristics of Mexico: a developing country with strong economic, historic, and geographic ties to the United States. In this paper I expand upon the language-earnings literature by examining offered wages in advertisements from Computrabajo.com.mx, a Mexican online job board.

Generally, internet job boards offer the advantage of being readily available and containing information that is not available in social surveys and other more traditional sources of information about the labor market (employers' requirements for English speakers in countries where English is not used extensively in business and/or government is a prime example of this). Furthermore, with some work these data can be compiled to provide almost immediate, very current information about firms, job seekers, and potentially the outcomes of their interaction in the labor market.

In this paper, to address issues of ability bias and measurement error typically found in the language-earnings literature, I exploit unique features of the job board data. I am able to include firm fixed effects, which controls for the possibility that high-wage firms disproportionately request English speakers (in which case the estimated return would be a firm effect, not an English effect). In addition, I control for a rich set of variables associated with cognitive and non-cognitive skills. Lastly, I argue that the specific requirement for an English speaker (or the lack of it) by a firm should reflect their actual need, and therefore might be more precise than self-reported bilingualism from survey data.

I see my estimates as complementary with the small number of person-fixed-effect estimates found in the literature. These studies have "tighter" ability controls, but they apply to a highly selected sample of persons who choose to learn English as a second language relatively late in life. At the same time, my access to firm fixed effects and to indicators for cognitive and non-cognitive skill demands improves on studies that rely on cross-sectional data.

The English premium I estimate is approximately 28 percent for the sample as a whole. This estimate is large, particularly considering that English is not an official language nor used extensively in government and business in Mexico. The estimated premium is equivalent

to the estimated wage premium associated with obtaining a college degree (compared to having only some college). Compared to the returns to English in other countries, my estimated premium is well above the 14 percent estimated for Israel by Lang and Siniver (2009) but in line with estimates for South Africa (27 percent estimated by Casale and Posel (2011*b*)) and India (34 percent estimated by Azam et al. (2010)).

I find no evidence of complementarity between language and other skills. Using education as an indicator for skill I find that jobs requiring higher levels of education show statistically lower returns to English language skills compared to jobs requiring a high school diploma or less. Using two alternative indicators for skill, experience and whether the job requires supervising or managing personnel, I find no difference in the English premia for advertisements with and without these requirements. Over all, this evidence suggests that speaking English may be an important path for upward economic mobility regardless of other skills, and it may actually be a more important one for those with low levels of other skills.

References

Angrist, J. D. and Lavy, V. (1997), 'The effect of a change in language of instruction on the returns to schooling in morocco', *Journal of Labor Economics* **15**(1).

Autor, D. H. (2001), 'Wiring the labor market', *The Journal of Economic Perspectives* **15**(1).

Azam, M., Chin, A. and Prakash, N. (2010), The returns to english-language skills in india. IZA Discussion Paper No. 4802. Available at SSRN: http://ssrn.com/abstract=1570400.

Bleakley, H. and Chin, A. (2004), 'Language skills and earnings: Evidence from childhood immigrants', *The Review of Economcs and Statistics* **86**(2).

Borjas, G. J. (1999), The economic analysis of immigration, *in* O. Ashenfelter and D. Card, eds, 'Handbook of Labor Economics', Vol. 3A, North-Holland, pp. 1697–1760.

Borjas, G. J. and Friedberg, R. M. (2009), 'Recent trends in the earnings of new immigrants to the united states', *NBER working paper series w15406* .

Brencic, V. (2011), 'Wage posting', *Canadian Journal of Economics* .

Brencic, V. and Norris, J. B. (2009), 'Employers' online search: An empirical analysis', *Industrial Relations* **48**(4).

Brencic, V. and Norris, J. B. (2010*a*), 'Do employers change job offers in their online job ads to facilitate search?', *Economic Letters* **108**(1).

Brencic, V. and Norris, J. B. (2010*b*), 'On-the-job tasks and performance pay: A vacancy-level analysis', *Industrial and Labor Relations Review* **63**(3).

Casale, D. and Posel, D. (2011*a*), 'English language proficiency and earnings in a developing country: The case of south africa', *The Journal of Socio-Economics* **40**(4).

Casale, D. and Posel, D. (2011*b*), 'Language proficiency and language policy in south africa: Findings from new data', *International Journal of Educational Development* **31**(5).

Chiswick, B. R. (1998), 'Hebrew language usage: Determinants and effects on earnings among immigrants in israel', *Journal of Population Economics* **11**(2).

Chiswick, B. S. and Miller, P. W. (1995), 'The endogeneity between language and earnings: International analysis', *Journal of Labor Economics* **13**(2).

Chiswick, B. S. and Miller, P. W. (2010), 'Occupational language requirements and the value of english in the us labor market', *Journal of Population Economics* **23**(1).

Christofides, L. N. and Swidinsky, R. (2010), 'The economic returns to the knowledge and use of a second official language: English in quebec and french in the rest-of-canada', *Canadian Public Policy-Analyse de Politiques* **36**(2).

Consulta Mitofsky (2007), `http://consulta.mx/web/` accessed on November 12, 2012.

Dustmann, C. (1994), 'Speaking fluency, writing fluency and earnings of migrants', *Journal of Population Economics* **7**(2).

Dustmann, C. and Van Soest, A. (2002), 'Language and the earnings of immigrants', *Industrial and Labor Relations Review* **55**(3).

Ginsburgh, V. A. and Prieto Rodriguez, J. (2011), 'Returns to foreign languages of native workers in the european union', *Industrial & Labor Relations Review* **64**(3).

Grenier, G. (1984), 'The effects of language characteristics on the wages of hispanic-american males', *The Journal of Human Resources* **19**(1).

Grenier, G. and Nadeau, S. (2011), English as the lingua franca and the economic value of other languages: the case of the language of work of immigrants and non-immigrants in the montreal labour market. working paper.

Hamilton, D., Goldsmith, A. H. and Darity Jr., W. (2008), 'Measuring the wage costs of limited english - issues with using interviewer versus self-reports in determining latino wages', *Hispanic Journal of Behavioral Sciences* **30**(3).

Instituto Nacional de Estadística y Geografía (2012), `http://www.inegi.org.mx` accessed on June 5, 2012.

Kuhn, P. and Shen, K. (2012), 'Do chinese employers avoid hiring overqualified workers? evidence from an internet job board', *Forthcoming, Research in Labor Economics* .

Kuhn, P. and Shen, K. (2013), 'Gender discrimination in job ads: Evidence from china', *Quarterly Journal of Economics* **128**.

Kuhn, P. and Skuterud, M. (2000), 'Job search methods: Internet versus traditional', *Monthly Labor Review* **123**(10).

LaLonde, R. J. and Topel, R. H. (1997), Economic impact of international migration and the economic performance of migrants, *in* M. R. Rosenzweig and O. Stark, eds, 'Handbook of Population and Family Economics. Handbooks in Economics 14', Vol. 1b, North-Holland, pp. 799–850.

Lang, K. and Siniver, E. (2009), 'The return to english in a non-english speaking country: Russian immigrants and native israelis in israel', *The B.E. Journal of Economic Analysis & Policy* **9**(1).

Marinescu, I. (2012), Online job search and unemployment insurance during the great recession. working paper.

Marinescu, I. and Wolthoff, R. (2012), Wages, applications, and skills. working paper.

McManus, W. S. (1990), 'Labor market effects of language enclaves: Hispanic men in the united states', *The Journal of Human Resources* **25**(2).

McManus, W. S., Gould, W. and Welch, F. (1983), 'Earnings of hispanic men: The role of english language proficiency', *Journal of Labor Economics* **1**(2), 101–130.

Mincer, J. (1974), *Schooling, experience, and earnings*, NBER Press, New York.

Morgan, S. L. and Winship, C. (2007), *Counterfactuals and Causal Inference (Analytical methods for social research)*, Cambridge University Press, Cambridge.

Munshi, K. and Rosenzweig, M. (2006), 'Traditional institutions meet the modern world: Caste, gender, and schooling choice in a globalizing economy', *American Economic Review* **96**(4).

Redstone Akrehs, I. and Frank, R. (2011), 'At the intersection of self and other: English language ability and immigrant labor market outcomes', *Social Science Research* **40**(5).

Saiz, A. and Zoido, E. (2005), 'Listening to what the world says: Bilingualism and earnings in the united states', *The Review of Economics and Statistics* **87**(3).

Secretaría de Economía, México (2012*a*), `http://www.economia.gob.mx` accessed on November 14, 2012.

Secretaría de Economía, México (2012*b*), `http://www.economia.gob.mx` accessed on November 14, 2012.

Secretaría de Educación Pública, México (2007), `http://promep.sep.gob.mx` accessed on January 13, 2013.

Stevenson, B. (2008), The internet and job search. NBER Working Paper No. 13886.

Tainer, E. (1988), 'English language proficiency and the determination of earnings among foreign-born men', *The Journal of Human Resources* **23**(1), 108–122.

United States Census Bureau (2012), `http://http://www.census.gov/foreign-trade/statistics/highlights/top/top1112yr.html` accessed on November 14, 2012.

A Tables

Table 1: Descriptive statistics, Computrabajo advertisements

Ad characteristic	
English requirements	
Requested English	0.088
Did not request English	0.912
Education requirements	
High school or less	0.554
Some postsecondary	0.150
University degree	0.296
Minimum experience requirements	
None required or less than one year	0.149
1-3 years	0.296
3-5 years	0.023
More than 5 years	0.005
No explicit experience requirement	0.527
Age requirements	
No age requirements	0.152
Ad specifies a minimum age	0.835
Ad specifies a maximum age	0.805
Mean age requested[1]	30.326
Wages	
Mean advertised salary[2]	$7,180
Other job characteristics	
Job is full time	0.891
Job requires supervision/management of personnel	0.069
Advertisement expresses male preference	0.160
Advertisement expresses female preference	0.165
Number of advertisements	10,840

1. Mean years of age requested. Includes only advertisements with a specified minimum and maximum age.

2. Mexican pesos per month. Includes only advertisements with a stated salary and is calculated as the average of minimum and maximum if a salary range is given. Excludes advertisements with salaries below 1,700 Mexican pesos per month and above 100,000 Mexican pesos per month

Table 2: Descriptive statistics by Computrabajo occupation

Computrabajo Occupation	Advertisement requested English	Average advertised salary (if advertised)[1]	Average age requested (if requested)[2]	Average minimum experience requested (if requested)[3]	Total observations
Administrative/Office	0.1025	$6,537	29.3	1.2	1,727
Art/Desgin/Media	0.1235	$7,799	28.9	1.2	81
Scientific/Research	0.1250	$9,019	30.8	1.5	56
IT/Telecomunications	0.2547	$10,339	29.2	1.3	530
Management	0.1241	$11,645	32.8	2.2	137
Economics/Accounting	0.1310	$8,343	30.2	1.3	687
Education/University	0.1891	$15,622	35.8	1.2	201
Lodging/Tourism	0.1469	$6,041	28.2	0.9	143
Engineering/Technical	0.2333	$9,443	30.1	1.5	840
Legal/Consulting	0.1556	$8,320	29.9	1.3	45
Marketing/Sales	0.0276	$6,445	31.0	1.0	3,259
Medicine/Health	0.0317	$8,252	30.4	1.1	315
Human resources	0.0926	$7,111	28.2	1.3	367
Other	0.0493	$5,740	30.3	1.1	2,452

1. Mexican pesos per month. Includes only advertisements with a stated salary and is calculated as the average of minimum and maximum if a salary range is given. Excludes advertisements with salaries below 1,700 Mexican pesos per month and above 100,000 Mexican pesos per month.

2. Mean years of age requested. Includes only advertisements with a specified minimum and maximum age.

3. Mean years of minimum experience requested. Includes only advertisements with a specified minimum experience.

Table 3: Descriptive statistics by education level, Computrabajo advertisements

Education level	Advertisement requested English	Average advertised salary[1]	Average age requested (if requested)[2]	Average minimum experience requested (if requested)[3]	Total observations
High school or less	0.0185	$5,393	30.5	0.9	6,006
Some postsecondary	0.0818	$6,788	29.5	1.3	1,626
University degree	0.2210	$10,722	30.4	1.6	3,208

1. Mexican pesos per month. Includes only advertisements with a stated salary and is calculated as the average of minimum and maximum if a salary range is given. Excludes advertisements with salaries below 1,700 Mexican pesos per month and above 100,000 Mexican pesos per month.

2. Mean years of age requested. Includes only advertisements with a specified minimum and maximum age.

3. Mean years of minimum experience requested. Includes only advertisements with a specified minimum experience.

Table 4: Comparison of descriptive statistics

	Advertised	Official Statistic
Salary		
Mean salary[1]	$7,180	$4,986
Age		
Mean age[2]	30	38
Education[3]		
Less than elementary	0.021	0.147
Elementary	0.000	0.208
Secondary	0.195	0.334
Medium-high education	0.784	0.311

1. Mexican pesos per month. Advertised includes only advertisements with a stated salary and is calculated as the average of minimum and maximum if a salary range is given. Excludes advertisements with salaries below 1,700 Mexican pesos per month and above 100,000 Mexican pesos per month. Average official salary is based on reports to the Mexican social security administrator (*Instituto Nacional de Estadística y Geografía*, 2012).
2. Mean years of age requested. Includes only advertisements with a specified minimum and maximum age. Official mean age is for the occupied population and is based on 2012 National Employment and Occupation data (*Instituto Nacional de Estadística y Geografía*, 2012).
3. Education shares are based on advertisements with explicit education requirements. Official education shares are based only on the population that indicated an education level and is based on 2012 National Employment and Occupation data (*Instituto Nacional de Estadística y Geografía*, 2012).

Table 5: Regression of log salary on English

VARIABLES	(1) ln_salary	(2) ln_salary	(3) ln_salary	(4) ln_salary	(5) ln_salary	(6) ln_salary	(7) ln_salary
English requested (1 if requested)	0.532***	0.273***	0.258***	0.251***	0.240***	0.253***	0.248***
	(0.0180)	(0.0160)	(0.0154)	(0.0154)	(0.0354)	(0.0359)	(0.0416)
Age		0.0720***	0.0689***	0.0747***	0.121***	0.118***	0.121***
		(0.0132)	(0.0131)	(0.0123)	(0.0240)	(0.0260)	(0.0322)
Age^2		-0.000586***	-0.000549**	-0.000718***	-0.00160***	-0.00160***	-0.00164***
		(0.000217)	(0.000216)	(0.000202)	(0.000385)	(0.000423)	(0.000532)
Minimum experience		0.132***	0.121***	0.126***	0.146***	0.144***	0.178***
		(0.00924)	(0.00927)	(0.00938)	(0.0134)	(0.0140)	(0.0349)
Minimum $experience^2$		-0.00298***	-0.00273***	-0.00293***	-0.00438***	-0.00425***	-0.0139
		(0.000347)	(0.000315)	(0.000307)	(0.000413)	(0.000409)	(0.00926)
Some postsecondary		0.226***	0.215***	0.187***	0.199***	0.216***	0.213***
		(0.0102)	(0.0101)	(0.0104)	(0.0191)	(0.0177)	(0.0154)
College degree		0.537***	0.516***	0.465***	0.445***	0.453***	0.463***
		(0.0101)	(0.0104)	(0.0107)	(0.0344)	(0.0376)	(0.0362)
Job requires supervising/managing personnel			0.0810***	0.0878***	0.137***	0.156***	0.173***
			(0.0152)	(0.0152)	(0.0248)	(0.0252)	(0.0227)
Cognitive and non-cognitive skill demands	NO	NO	YES	YES	YES	YES	YES
State fixed effects	NO	NO	YES	YES	YES	YES	YES
Occupation fixed effects	NO	NO	NO	YES	YES	YES	YES
Company fixed effects	NO	NO	NO	NO	YES	YES	YES
Company*Occupation fixed effects	NO	NO	NO	NO	NO	YES	YES
Company*Occupation*State fixed effects	NO	NO	NO	NO	NO	NO	YES
Adjusted R-squared	0.073	0.404	0.446	0.470	0.596	0.620	0.640
Observations	10,840	10,840	10,840	10,840	10,840	10,840	10,840

Robust standard errors in parentheses. Standard errors clustered by company in columns 5 through 7. *** p<0.01, ** p<0.05, * p<0.1

Baseline case for education categories are jobs explicitly indicating an education requirement of high school or below. The regression in column 2 also includes controls for advertised gender preference and for whether age and/or experience were missing. Columns 3 through 7 include a control for whether the job is advertised as full-time employment.

Table 6: Regression output, log of wages on English and interactions with education requirements, supervision/management requirements, and minimum experience requirements

VARIABLES	(1) ln_salary	(2) ln_salary	(3) ln_salary	(4) ln_salary	(5) ln_salary
English requested (1 if requested)	0.248***	0.401***	0.243***	0.256***	0.412***
	(0.0416)	(0.0764)	(0.0405)	(0.0519)	(0.0830)
Minimum experience	0.178***	0.176***	0.178***	0.183***	0.182***
	(0.0349)	(0.0352)	(0.0347)	(0.0369)	(0.0365)
Minimum experience^2	-0.0139	-0.0136	-0.0139	-0.0153	-0.0153
	(0.00926)	(0.00941)	(0.00923)	(0.0100)	(0.0100)
Some postsecondary	0.213***	0.218***	0.213***	0.213***	0.218***
	(0.0154)	(0.0181)	(0.0158)	(0.0154)	(0.0183)
College degree	0.463***	0.477***	0.464***	0.463***	0.478***
	(0.0362)	(0.0390)	(0.0366)	(0.0360)	(0.0391)
Job requires supervising/managing personnel	0.173***	0.170***	0.165***	0.172***	0.160***
	(0.0227)	(0.0227)	(0.0214)	(0.0226)	(0.0218)
English*Some postsecondary		-0.159**			-0.164**
		(0.0771)			(0.0746)
English*College degree		-0.187***			-0.196***
		(0.0589)			(0.0568)
English*Job requires supervising/managing personnel			0.0984		0.109
			(0.0869)		(0.0879)
English*Minimum experience				-0.0364	-0.0452
				(0.0436)	(0.0395)
English*Minimum experience^2				0.0102	0.0125
				(0.0120)	(0.0112)
Adjusted R-squared	0.640	0.641	0.641	0.640	0.642
Observations	10,840	10,840	10,840	10,840	10,840

Standard errors clustered by company. *** p<0.01, ** p<0.05, * p<0.1

All regressions include age and age squared, controls for cognitive and non-cognitive skill demands, controls for whether the job is advertised as full-time, controls for gender preference, and for whether age and/or experience are missing. Baseline case for education categories are jobs explicitly indicating an education requirement of high school or below. All specifications also include Company*Occupation*State fixed effects.

Table 7: Regression output, log of wages on covariates, propensity-score-matched sample

VARIABLES	(1) Main specification	(2) Propensity-score matches	(3) Main interacted specification	(4) Propensity-score matches
English requested (1 if requested)	0.248***	0.235***	0.412***	0.369***
	(0.0416)	(0.0297)	(0.0830)	(0.0511)
Minimum experience	0.178***	0.120**	0.182***	0.102
	(0.0349)	(0.0516)	(0.0365)	(0.0648)
Minimum experience^2	-0.0139	0.00303	-0.0153	0.00138
	(0.00926)	(0.00851)	(0.0100)	(0.0122)
Some postsecondary	0.213***	0.0354	0.218***	0.145**
	(0.0154)	(0.0410)	(0.0183)	(0.0572)
College degree	0.463***	0.339***	0.478***	0.469***
	(0.0362)	(0.0417)	(0.0391)	(0.0542)
Job requires supervising/managing personnel	0.173***	0.112**	0.160***	0.0656
	(0.0227)	(0.0522)	(0.0218)	(0.0915)
English*Some postsecondary			-0.164**	-0.171**
			(0.0746)	(0.0705)
English*College degree			-0.196***	-0.197***
			(0.0568)	(0.0581)
English*Job requires supervising/managing personnel			0.109	0.0763
			(0.0879)	(0.120)
English*Minimum experience			-0.0452	0.0192
			(0.0395)	(0.0500)
English*Minimum experience^2			0.0125	0.00353
			(0.0112)	(0.0116)
Adjusted R-squared	0.640	0.597	0.642	0.600
Observations	10,840	4,549	10,840	4,549

Standard errors clustered by company. *** p<0.01, ** p<0.05, * p<0.1

All regressions include age and age squared, controls for cognitive and non-cognitive skill demands, controls for whether the job is advertised as full-time, controls for gender preference, and for whether age and/or experience are missing. Baseline case for education categories are jobs explicitly indicating an education requirement of high school or below. All specifications also include Company*Occupation*State fixed effects.

B Figures

Figure 1: Predictive margins and 95 percent confidence intervals, education requirement

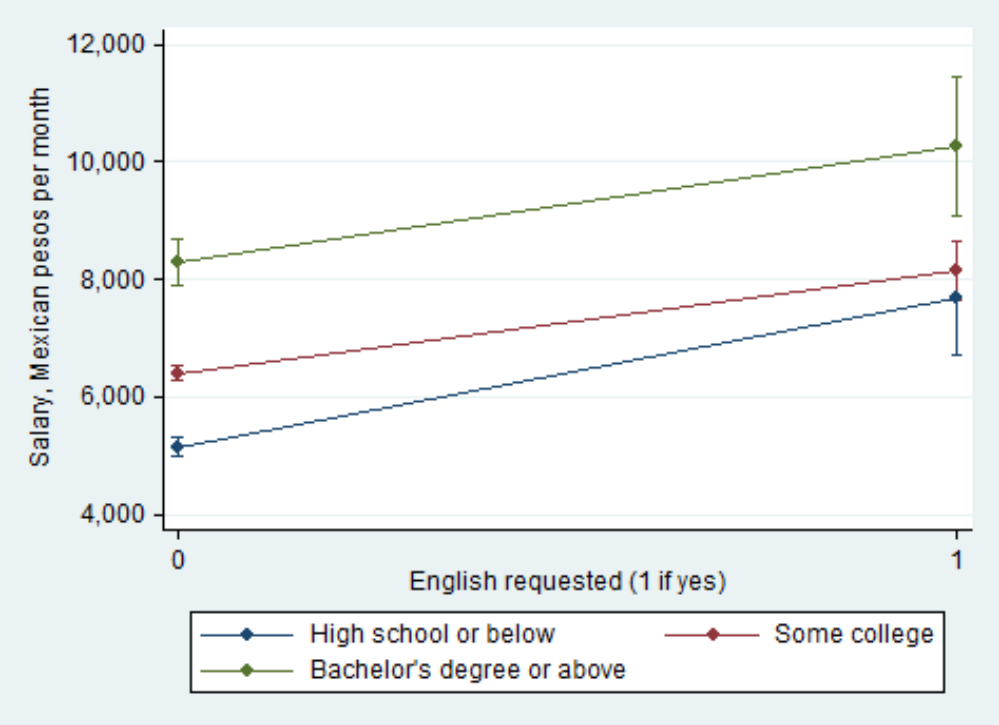


Figure 2: Box plots, comparison of advertised salaries

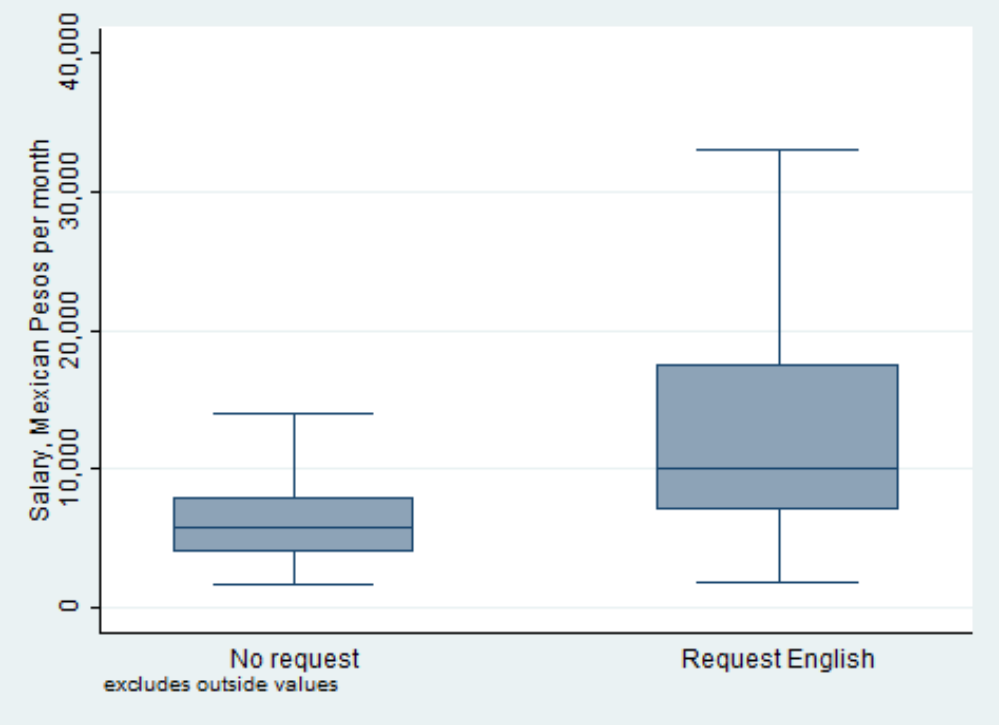
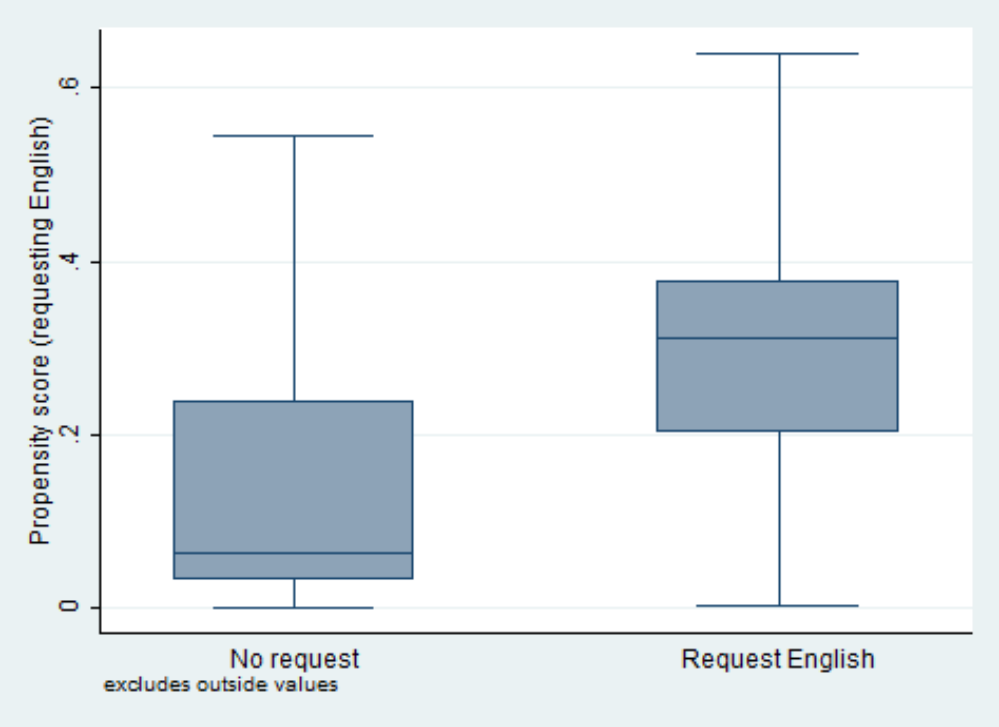
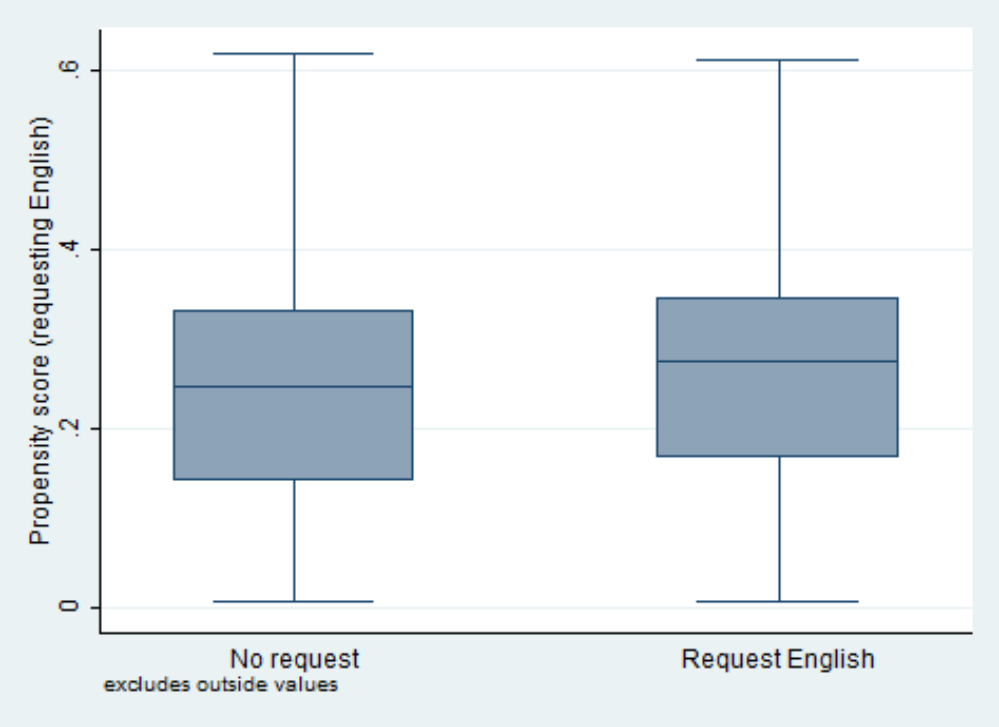


Figure 3: Box plots, comparison of propensity scores

(a) Propensity scores, whole sample



(b) Propensity scores, matched sample only



C Data sources, collection, and sample

This study uses data from job advertisements posted for employment in Mexico on Computrabajo.com, an online job board with job postings in Latin American countries and Spain. The data I analyze consist of unique job advertisements posted on the site's Mexican extension, Computrabajo.com.mx, where unique advertisements include advertisements that appeared once during the data collection time frame, as well as identical advertisements appearing on multiple dates (in which case only the first appearance is included in the sample). Using a combination of the uniform resource locator (url) and the occupation title from each advertisement duplicate observations as described above were removed.

The advertisements on Computrabajo can be thought of consisting of two parts; one part with data fields, and one part with text. The fields in the first part include the advertisement's date of posting and the occupation, as well as application instructions, the name of the company, a company code, a job type (full-time versus part-time, work-from-home, and others) field, a job period (permanent, undetermined, and others) field, a start date field, a salary field, and a state field. Some of these fields are selectable (such as job category and state), whereas several are open fields (such as job type, job period, and salary). The second part of the advertisement is an open text field where the potential employers may write additional job description, compensation, requirements, or essentially any other information they want to provide in the advertisement.

On a daily basis, after all data were collected, each advertisement was loaded onto a relational database, generating two tables: a table assigning a unique identifier to text fields (referred to as text table), and a table with advertisement characteristics obtained from the fields described above (referred to as field table), with the unique identifier serving as the link between the two tables. As stated above, advertisements could be posted under different occupations, which would generate multiple entries in the field table but not in the text table. For example, an advertisement appearing in two occupations the same day would have two entries in the field table (one for each occupation), but since the text that generates the unique identifier is unchanged, only one entry appears in the text table. A particular advertisement could be repeated in more ways than appearing in multiple categories. The same advertisement could be posted on different dates, or posted with a different salary, start date, or any of the field entries (as well as combinations of these such as re-posting on a different date, under a different category, and with a different salary), as long as the text portion of the advertisement remained unchanged.

For this analysis, any observations that were repeated across dimensions other than time were eliminated. During the approximately 18-month period from January 25 2011 through

August 25 2012, a total of 1,607,972 advertisements were posted. When sub-sampling to first appearances of each posting (using the url-job title-state combination), my sample is reduced to 202,748 observations.

The variables used in the analysis were searched for in three different ways. For variables in the field table that are selectable, such as state, the values were obtained directly from the fields. For variables in the field table that are open fields, some interpretation and coding was necessary; For example, the salary field was an open field so the text entered was interpreted and translated to numerical values. Lastly, variables obtained from the text part of the advertisement were obtained by parsing the text searching for key words, and then interpreting and coding the text following the search. For example, the words age and *edad* (age in Spanish) were searched for, and the text following these words was captured, interpreted as numerical values, and coded accordingly. Following are more detailed descriptions of how each variable was collected.

The age variable was obtained by parsing the text searching for the words age and *edad*, which returned a list of entries containing the text that followed these key words for each advertisement that contained these words. Duplicates were removed from the list and then four variables were created for each unique line of text. These variables are minimum age requested, maximum age requested, missing minimum age (equal to 1 if no minimum was requested), and missing maximum age (equal to 1 if no maximum was requested). Notice that words containing age (such as manage) or *edad* (such as *notoriedad*) generated an entry on the list, but the text returned in these cases did not provide age information, so these entries were coded as 1 in both the missing minimum age and the missing maximum age variables. Once these variables were created, each observation was matched using the text to create these variables in the main dataset. For observations without a match, the age variables were coded as zero and the missing age variables were coded as one.

The occupation was coded directly from the job category that the advertisement was filed under. Given that I eliminated observations that were posted across multiple occupations, each advertisement in the data analyzed has only one occupation associated to it. The company and the state for which the advertisement is for were also coded directly from the company and state fields.

The education variable was obtained by parsing the text searching for the words educ (for education and *educación*) and *escolaridad* (schooling in Spanish), which returned a list of entries containing the text that followed these key words for each advertisement that contained these words. Duplicates were removed from the list and then one variable was created and coded as a specific education level for each unique line of text.

Some assumptions were necessary when making this classification, as not all advertise-

ments state clearly the required education. Many advertisements state a minimum requirement but may include a desired level of education that is higher than that minimum. In all these instances I selected the minimum requirement. Also, many advertisements indicated desired levels of education such as university or college, or like engineering or law. In such instances I coded them as "some postsecondary". When advertisements indicated such education levels but also included words such as completed or with a degree I coded them as "college degree". Also, when the requirement was for a title, such as Engineer, rather than a career or degree, such as engineering, I also coded them as "college degree".

Technical careers (*Carrera Tecnica*) were coded as "Technical degree". A word that appeared repeatedly is *Trunco*, which is used to mean interrupted or unfinished, so an education requirement using the word *trunco* was interpreted to mean partial completion of, for example, a high school diploma or a university degree, being coded, respectively, as "less than high school" and "Some postsecondary". The words *Bachillerato*, and *Preparatoria* were interpreted as "high school", whereas *secundaria* was treated as equivalent to middle school and therefore coded as "less than high school". Due to small sample sizes of some particular education levels, in the analysis education levels were aggregated into "high school or below", "some postsecondary", and "college degree".

The English variables were obtained by parsing the text searching for the words English, *Ingl* (for *Inglés*, the Spanish word for English. We used only the first four letters to capture misspelled versions of the word), and *Biling* to capture both bilingual and *Bilingüe* (the Spanish word for bilingual). An indicator variable *english* was then generated, populated with one if an advertisement contained any of these key words and a zero otherwise.

The experience variable was obtained by parsing the text searching for the word exp, which returned a list of entries containing the text that followed this key word for each advertisement that contained it. Duplicates were removed from the list and then four variables were created for each unique line of text. These variables are minimum experience requested, maximum experience requested, missing minimum experience (equal to 1 if no minimum was requested), and missing maximum experience (equal to 1 if no maximum was requested). Once these variables were created, each observation was matched using the text to create these variables in the main dataset. For observations without a match, the age variables were coded as zero and the missing experience variables were coded as one. In the analysis I use minimum experience requests only.

Gender variables were obtained by parsing the text searching for the words sex, gender, and *género*, which returned a list of entries containing the text that followed these key words for each advertisement that contained them. Duplicates were removed from the list and then three indicator variables were created for each unique line of text. These variables are male

requested, female requested, and no gender preference. Once these variables were created, each observation was matched using the text to create these variables in the main dataset.

The salary variable was obtained from the advertisements' salary field. This field, however, is an open field, not a selection box. The salary text was collected from all advertisements and after removing duplicates a list of unique salary texts was created. Based on the text, the following salary variables were created; salary minimum, salary maximum, salary periodicity, and salary currency. Additionally, indicator variables were created for whether the salary included extras such as commissions.

Using the salary text field from each observation, the salary fields were added to each observation, and then all observations were transformed to monthly salaries in Mexican pesos. Advertisements in many Latinamerican countries tend to indicate a monthly salary. For advertisements where it is not otherwise indicated, we assume that the advertised salary is monthly. Advertisements using piece wages (with the exception of hourly, non-output-dependent) generate missing salary information. All other wages are converted assuming a monthly salary. That is, advertisements with hourly wages use a monthly salary of 160 times that hourly rate. Similarly, daily rates are converted to monthly by multiplying by 20, weekly salaries are converted by multiplying by 4, fortnightly salaries are converted by multiplying by 2, and annual salaries are converted by dividing by 12.

Most advertisements are assumed to be indicating a salary in Mexican pesos, unless clearly stated otherwise. In the instances when Mexican pesos are not indicated, U.S. dollars are usually the currency used. In some cases it is clearly indicated, such as USD, in other cases it is inferred from context, such as in advertisements in English using an annual salary that appears in line with what should be U.S. dollars. A few advertisements state Canadian dollars or Euros as the currency. A conversion was made to Mexican pesos from all these currencies using exchange rates from the day of the advertisement.

Advertised salaries were determined as the average of the minimum and maximum advertised salary for each advertisement, where the minimum and the maximum were the same if a salary rather than a range was advertised. Advertisements with monthly salaries equivalent to 1,700 Mexican pesos or less were dropped from the analysis since these would be essentially below minimum wage. Advertisements with salaries above 100,000 Mexican pesos per month were also dropped, as were advertisements where the salary range (maximum-minimum) was more than double the minimum.

Cognitive and non-cognitive skill demand variables were created by parsing the text of each advertisement searching for key words associated to each skill and populating each skill variable with a one if an associated key word was found in the advertisement and a zero otherwise. The non-cognitive skill demands are: ambitiousness, assertiveness, com-

mitment, communication skills, compliance, creativity, credit worthiness, detail oriented, discipline, discreet, dynamism, emotional stability, extroversion, flexibility, goal oriented, honesty, independence, initiative, interpersonal skills, leadership skills, negotiation skills, orderly, passion, patience, positive attitude, punctuality, resilience, respect, responsibility, self motivation, sense of urgency, service oriented, teamwork oriented, and (ability to) work under pressure. Cognitive skill demands are: analytical, fast learner, numerical ability, and problem solver.

The main sample I used for the analysis represents only a portion of all advertisements collected. Table C.1 presents a comparison of descriptive statistics for different samples of the data. The share of advertisements that request English speakers decreases with the imposition of each sample restriction, from slightly less than 20 percent for all unique observations to just under 9 percent for the main sample analyzed. Similarly, the share of advertisements with education requirements of high school or below generally increases as sample restrictions are imposed.

Observations with explicit education requirements are more likely to specify age and experience requirements, as well as a gender preference, than all advertisements. This is also true for advertisements with a stated salary, but to a smaller extent. Average advertised salaries are about 500 Mexican pesos per month lower when education requirements are explicit, and an additional 500 Mexican pesos per month lower when further sampling to observations from non-staffing firms (last column). The average age requested, as well as the share of jobs advertised as full time and the share of jobs requiring supervision or management of personnel do not vary significantly as sample restrictions are imposed.

In general imposing my sample restrictions appears to lead to selection of less-skilled jobs than the average advertised job on Computrabajo.com. While I control for observable skills, if there are complementarities between language and skill, I will underestimate the English premium in Computrabajo-advertised jobs as a whole as a result of imposing my sample restrictions.

Table C.1: Comparison of descriptive statistics by sample criteria

Ad characteristic	All unique observations	All observations with education requirements	All observations with salaries	All observations with education requirements and salaries	Main sample
English requirements					
Requested English	0.198	0.164	0.137	0.121	0.088
Did not request English	0.802	0.836	0.863	0.879	0.912
Education requirements					
High school or less	0.190	0.429	0.250	0.496	0.554
Some postsecondary	0.067	0.151	0.085	0.168	0.150
University degree	0.186	0.421	0.169	0.336	0.296
No explicit education requirement	0.557	NA	0.497	NA	NA
Minimum experience requirements					
None required or less than one year	0.073	0.103	0.096	0.119	0.149
1-3 years	0.207	0.283	0.224	0.297	0.296
3-5 years	0.052	0.061	0.037	0.040	0.023
More than 5 years	0.016	0.017	0.011	0.010	0.005
No explicit experience requirement	0.652	0.537	0.633	0.533	0.527
Age requirements					
No age requirements	0.478	0.214	0.407	0.182	0.152
Ad specifies a minimum age	0.508	0.768	0.579	0.802	0.835
Ad specifies a maximum age	0.491	0.742	0.562	0.777	0.805
Mean age requested[1]	30.653	30.658	30.179	30.203	30.326
Wages					
Mean advertised salary[2]	$8,104	$7,652	$8,105	$7,652	$7,180
Other job characteristics					
Job is full time	0.893	0.916	0.887	0.913	0.891
Job requires supervision/management of personnel	0.062	0.070	0.059	0.062	0.069
Advertisement expresses male preference	0.109	0.161	0.135	0.186	0.160
Advertisement expresses female preference	0.114	0.159	0.139	0.179	0.165
Number of advertisements	202,748	89,874	49,100	24,721	10,840

1. Mean years of age requested. Includes only advertisements with a specified minimum and maximum age.

2. Mexican pesos per month. Includes only advertisements with a stated salary and is calculated as the average of minimum and maximum if a salary range is given. Excludes advertisements with salaries below 1,700 Mexican pesos per month and above 100,000 Mexican pesos per month.