J Popul Econ (2010) 23:353–372
DOI 10.1007/s00148-008-0230-7

ORIGINAL PAPER

Occupational language requirements and the value of English in the US labor market

Barry R. Chiswick · Paul W. Miller

Received: 8 March 2007 / Accepted: 30 October 2008 /
Published online: 14 January 2009

Abstract This paper is concerned with the English-language requirements of occupations in the USA, as measured by the O*NET database. These scores are linked to employed native and foreign-born men in the 2000 Census. Earnings increase with the respondent's proficiency in English, with the English proficiency required for the occupation, and when those with high levels of proficiency work in jobs requiring English-language skills (interaction effect). There is a strong economic incentive for the matching of worker's English skills and the occupation's requirements, and this matching tends to occur in the labor market.

Keywords English language · Earnings · Occupation

JEL Classification J240 · J310 · J620

Responsible editor: Klaus F. Zimmermann

B. R. Chiswick (✉)
Department of Economics, University of Illinois at Chicago, 601 South Morgan Street, Chicago, IL 60607-7121, USA
e-mail: brchis@uic.edu

B. R. Chiswick
IZA-Institute for the Study of Labor, Bonn, Germany

P. W. Miller
Business School, University of Western Australia, Perth, WA, Australia

1 Introduction

A focus of a number of recent studies of the US labor market has been on whether individuals who are more proficient in English have higher earnings than those less proficient in English-language skills. Studies for immigrants have generally shown that proficiency in English is associated with a substantial earnings premium (e.g., Chiswick and Miller 1995). Studies for the native born, however, have shown that English–other language bilingualism may actually be associated with lower earnings (e.g., Fry and Lowell 2003; Chiswick and Miller 2007).

The current study seeks to establish whether this earnings premium for immigrants who are proficient in English in the USA arises because the immigrants have higher earnings within an occupation or whether they do better by being able to gain entry to occupations that have higher English-language requirements. The issue of the lower earnings of the native born who speak a non-English language compared to those who at home are monolingual English speakers is addressed within the same framework. The research thus builds on earlier work which argued that occupation was the intervening activity that linked earnings to human capital skills (e.g., Groshen 1991; Duncan 1961). Duncan (1961, pp.116–117), for example, stated that "…a man qualifies himself for occupational life by obtaining an education; as a consequence of pursuing his occupation, he obtains income. Occupation, therefore, is the intervening activity linking income to education." In other words, education, or any other human capital skill, such as English-language proficiency, can be viewed as having both direct and indirect impacts on earnings, with the indirect impacts operating via occupational attainment.

This study uses a range of complementary techniques to explore the links between English-language skills and earnings. The baseline methodology is an encompassing specification of a human capital earnings equation that contains variables for both the individuals' English-language skills and the English-language requirements of occupations. An instrumental variables (IV) model is also employed, wherein the individuals' English-language skills affect occupational attainment (indexed by English-language requirements), and the English-language requirements of the job in turn affect earnings. Finally, the robustness of the findings is examined via the two-step procedure for analysis of samples that combine aggregate (in this instance, occupation-level) data with micro-level data proposed by Dickens and Katz (1987).

Section 2 provides an overview of the model, the Census data, and the information on the English-language requirements of jobs in the US labor market utilized in the analysis. Section 3 presents the findings of the statistical analyses. These analyses are conducted separately for native-born men and for foreign-born men. Section 4 provides concluding comments.

2 Model and data overview

2.1 Methodological issues

The analyses presented below are based on an earnings equation that contains information on the respondent's self-reported proficiency in English (PROF_i), a set of other human capital and demographic standardizing variables (X_i), and the English requirements of the occupation in which the individual was employed (ER_i). This model may be written as:

$$\ln Y_i = \alpha + \text{PROF}_i\beta + X_i\gamma + \text{ER}_i\delta + \varepsilon_i, \tag{1}$$

where income (Y) is a measure of annual earnings from wage and salaried employment or self-employment, and ε_i is a stochastic disturbance term. The X_i vector includes educational attainment, potential labor market experience, location, race, marital status, birthplace, and among the foreign born, duration of residence in the USA.

This benchmark model can be extended in various ways. One is to include occupational fixed effects in the analysis. As the ER_i variable is constructed with reference to the individual's occupation, the occupation fixed effects can only be controlled for at a different level of aggregation. The approach taken below is to construct the ER_i variable using information on the 509 three-digit occupations in the 2000 Census and to control for the occupation fixed effects using the 23 broad Census occupational categories. A second extension involves adding to the equation interaction terms between the ER_i variable and the workers' self-reported proficiency in English (PROF_i). These interaction terms will indicate if there is a return to correct matching of workers to the English-language requirements of occupations. A third extension is an examination of whether the relationship between earnings and the English-language requirements of occupations is nonlinear.

The ER_i variable in Eq. 1 may be endogeneously determined. The standard response to this is the use of instrumental variables estimation. A difficulty with this approach is the limited number of instruments in the data set.[1] An assessment of the potential importance of endogeneity is provided, however, in models that use the individual's proficiency in English (PROF_i) and functional

[1] In the words of Heckman et al. (2003), "the Census data used in this paper yield large samples but few instruments."

form for identification.[2] Thus this approach may be viewed as estimation of the following equation:

$$\ln Y_i = \alpha + X_i\gamma + \widehat{\mathrm{ER}}_i\delta + \varepsilon_i, \text{ where} \tag{2a}$$

$$\widehat{ER}_i = f(\text{level of education}_i) + Z_i\widehat{\varphi} + \mathrm{PROF}_i\widehat{\lambda}. \tag{2b}$$

This approach to identification follows the practice of some studies in the sample selection literature, where different representations of educational attainment and age/experience have been used in the selection equation and the wage equation. For example, education has been entered in the selection equation as dummies for the various levels of achievement, and it has then been entered in the wage equation as "years of education" (see, for example, Hartog and Oosterbeek 1993; Gyourko and Tracy 1988). See also the discussion and use of this strategy in IV estimation in Meng and Gregory (2005, p.157).

A further potential problem with the analysis above is that it combines variables that are measured at two different levels of aggregation. Thus all data other than that for English requirements are individual-level data, whereas the ER information is grouped data, constructed at the level of the occupation. Dickens and Katz (1987) and Moulton (1986) draw attention to the possibility that combining aggregate data with individual data may bias the estimates and also lead to incorrect standard errors.

Dickens and Katz (1987) outline an expedient approach that can be followed in the current analysis to check the reliability of the results in the face of this potential problem. They propose that the earnings equation in Eq. 1 be estimated without the ER_i variable but allowing for detailed occupation fixed effects (i.e., for the 509 three-digit Census occupations), namely:

$$\ln Y_i = \alpha + X_i\gamma + \mathrm{PROF}_i\delta + \mathrm{OCC}_i\phi + \varepsilon_i. \tag{3a}$$

The occupation fixed effects (i.e., the $\widehat{\phi}$) are then regressed on the O*NET English requirements variables in a subsequent regression, estimated at the level of the occupation (j).

$$\widehat{\phi}_j = \alpha_0 + \alpha_1\mathrm{ER}_j + \eta_j. \tag{3b}$$

This two-step approach should provide reliable estimates of the links between earnings and the occupational English requirements. It also provides information on the extent to which the across-occupation variation in earnings (after standardization for the workers' characteristics) can be accounted for using information on the occupational English requirements.

[2]The use of PROF_i in this regard follows from suggestions by a referee. It is consistent with the overall approach of the study.

2.2 Data

Information on earnings, the human capital and demographic variables, as well as the self-reported proficiency in English is obtained from the 2000 US Census Public Use MicroData Sample (1% sample). The focus is on men aged 25 to 64 years who worked in paid employment in 1999. All variables are described in detail in Appendix A. The novel feature of this study is the inclusion in the estimating equation of information on the English requirements of occupations.

Information on the English requirements of jobs is obtained from the Occupational Information Network, or O*NET, database.[3] The September 2007 release, O*NET version 12.0, is used. This contains two sets of information relevant to the analysis. The first is about "How *important* is knowledge of the ENGLISH LANGUAGE to the performance of *your current job*?" (emphasis in O*NET Knowledge Questionnaire). The second is "What *level* of ENGLISH LANGUAGE is needed to perform *your current job*?" (emphasis in original).

The information on the *importance* of the English language was collected on a five-point scale: (1) Not important; (2) Somewhat important; (3) Important; (4) Very important; and (5) Extremely important. The information on the *level* of English-language proficiency needed to perform in the current job was collected only among those who felt that English was somewhat or more important to the performance of their current job. A seven-point scale was used, with three benchmark descriptors offered as a guide: 2 = write a thank you note; 4 = edit a feature article in a local newspaper; and 6 = teach a college English class. Individuals who did not feel that English was important to the performance of their current job were coded as zero on the scale for level of English. Hence, it is an eight-point scale (0–7).

To make the O*NET data more intuitively understandable to users, descriptor average ratings were standardized to a scale ranging from 0 to 100. This is accomplished using the formula:

$$S = \big((O - L)/(H - L)\big) \times 100 \tag{4}$$

where S is the standardized score, O is the original rating score, L is the lowest possible score on the rating scale used, and H is the highest possible score on the rating scale used. The standardization procedure assumes an equal distance between points on the five-point and eight-point scales. Hence, the original scores on the five-point importance of English scale become 1 = 0; 2 = 25;

[3]The National O*NET Consortium was organized to develop O*NET and its related products for the US Department of Labor Employment and Training Administration (ETA). The Consortium currently comprises the O*NET Management Partnership (the North Carolina Employment Security Commission, the Center for Employment Security Education and Research, and the ETA) and the National O*NET Support Group (composed primarily of the National Center for O*NET Development, the Research Triangle Institute, and the Human Resources Research Organization). Web address of O*NET data: http://online.onetcenter.org.

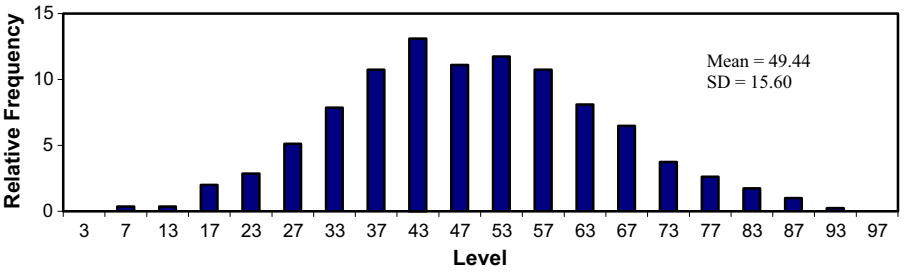


Fig. 1 Relative frequency of occupations on standardized score of level of English

3 = 50; 4 = 75; and 5 = 100. The scores on the eight-point level of English scale become 0 = 0; 1 = 14.3; 2 = 28.6; 3 = 42.9; 4 = 57.1; 5 = 71.4; 6 = 85.7; 7 = 100. The data on both the importance and level of English are presented in the form of average scores for specific occupations in the O*NET database.

There is a very high correlation (0.92) between the scores for the importance of English and the level of English. That is, occupations where knowledge of the English language is held to be important to job performance are occupations where a relatively high level of English-language proficiency is needed to perform in the job. This carries across to analysis within each broad major occupational category, where the correlation coefficient between the two O*NET measures ranges from a low of 0.82 for "Service Occupations" to a high of 0.90 for the "Construction, Extraction, and Maintenance Occupations." While parallel analyses were conducted using the importance of English and level of English information, the findings were, as expected given these high correlations, very similar. For space reasons, only the results for the level of English data are presented.[4]

The data on the level of English for all occupations are illustrated in Fig. 1, with data for a high-requirement occupational category (Management, Professional, and Related Occupations) and for a low-requirement occupational category (Production, Transportation, and Material Moving Occupations) being presented in Fig. 2.

The mean for all occupations is 49.4, which is two thirds of the way between levels 29 and 57 which had the benchmark descriptors of "write a thank you note" and "edit a feature article in a local newspaper." The standard deviation of the standardized score is 16. The occupations in the US labor market cover a wide range of the standardized scores, although there is limited representation above scores of 80 and below scores of 20.

Examination of the standardized scores for the level of English needed to perform the job in the selected occupational categories shows that high-skilled occupations, such as "Management, Professional and Related Occupations",

[4]Findings from the analyses based on the importance of English are presented in a Statistical Appendix, available from the authors upon request.

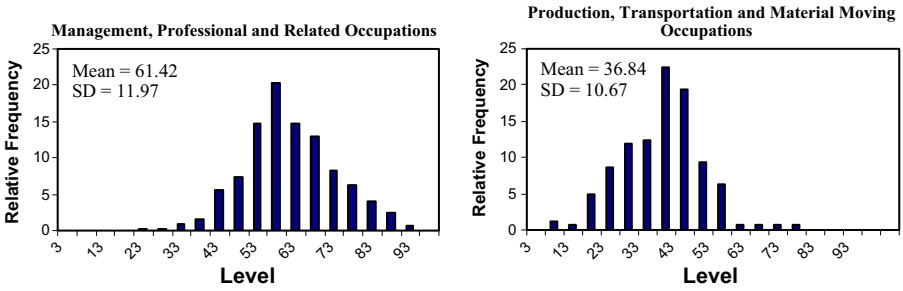


Fig. 2 Relative frequency of occupations on standardized score of level of English

have a high required level of English and a distribution that is slightly right-skewed (left-hand side of Fig. 2). Low-skilled occupations, such as "Production, Transportation, and Material Moving Occupations", have a low required level of English and have a distribution of scores skewed towards the left (right-hand side of Fig. 2).

The Census occupation codes (509 separate occupations) and the Standard Occupational Classification System (SOC) codes (812 separate occupations) in the O*NET database do not always correspond, and so a number of approximations are required when linking the two data sets.[5] Where the codes exactly match the O*NET score is assigned to the Census occupational category. Where English requirements were provided in the O*NET database only for subgroups of a Census occupational category, the simple average of the scores for these subgroups was used. In a small number of cases, data on the English requirements were not available for Census occupations, and the score for a similar occupation was used. Finally, averages of occupations in the relevant categories were used for the Census "all other" categories.

Following the assignment of the occupational English scores outlined above, the mean level of English required for their jobs for native-born male workers is 48.3 and that for foreign-born male workers is about 5% lower, at 45.7. The analyses which follow show how these English-language requirements impact the individual's earnings.

3 English-language requirements and earnings

3.1 Benchmark model and simple extensions

The analysis of the determinants of earnings first reports results for the benchmark model of Eq. 1. Then findings from the first two extensions

[5]The O*NET database version 12.0 contains a crosswalk between the SOC that forms the basis of the Census codes and the O*NET codes. Three hundred sixty-nine of the 509 Census 2000 occupational categories exactly match SOC categories, 127 match at a broader level, and 13 are aggregates of SOC categories.

Table 1 Earnings functions with focus on level of English skill, by nativity, 2000 US Census

Variable	Native born			Foreign born		
	(i)	(ii)	(iii)	(i)	(ii)	(iii)
Constant	4.249 (190.55)	4.426 (187.72)	4.244 (190.02)	5.307 (97.47)	5.654 (101.20)	5.060 (84.16)
Log weeks worked	1.002 (182.88)	0.980 (180.26)	1.002 (182.91)	0.870 (73.02)	0.854 (72.25)	0.868 (73.00)
Years of education	0.071 (114.90)	0.073 (113.04)	0.071 (114.94)	0.035 (44.27)	0.030 (35.53)	0.035 (43.04)
Experience	0.035 (76.68)	0.033 (73.89)	0.035 (76.65)	0.015 (14.84)	0.018 (17.59)	0.016 (15.46)
Experience squared/100	−0.061 (61.34)	−0.057 (58.00)	−0.061 (61.32)	−0.024 (12.30)	−0.028 (14.78)	−0.025 (12.91)
Married	0.256 (106.69)	0.232 (98.19)	0.256 (106.68)	0.201 (34.38)	0.179 (31.39)	0.198 (34.01)
South	−0.063 (27.87)	−0.068 (30.45)	−0.063 (27.87)	−0.078 (13.18)	−0.086 (14.79)	−0.078 (13.22)
Metropolitan area	0.193 (32.82)	0.188 (32.18)	0.193 (32.80)	0.126 (4.73)	0.097 (3.70)	0.131 (4.92)
Veteran	−0.037 (13.56)	−0.041 (15.22)	−0.036 (13.53)	−0.061 (4.76)	−0.047 (3.70)	−0.054 (4.21)
Blacks	−0.135 (38.14)	−0.096 (27.29)	−0.135 (38.03)	−0.161 (15.17)	−0.106 (10.19)	−0.153 (14.36)
Years since migration (YSM)				0.012 (14.94)	0.012 (15.26)	0.012 (14.83)
YSM squared/100				−0.014 (7.37)	−0.014 (7.14)	−0.014 (7.44)
English very well	−0.066 (13.15)	−0.048 (9.69)	0.018 (0.85)	−0.071 (7.84)	−0.070 (7.89)	−0.015 (0.40)
English well	−0.106 (8.36)	−0.089 (7.07)	−0.093 (1.87)	−0.200 (20.35)	−0.169 (17.61)	0.239 (6.40)
English not well/not at all	−0.010 (0.52)	0.008 (0.42)	0.069 (0.94)	−0.298 (28.27)	−0.267 (25.76)	0.295 (7.55)
Level of English[a]	1.288 (112.51)	1.337 (67.88)	(a)	1.553 (56.54)	1.107 (26.40)	(a)
Level/*English only/			1.297 (111.26)			2.075 (33.44)
Level*English very well			1.123 (26.47)			1.962 (46.86)
Level*English well			1.269 (12.48)			1.145 (22.87)
Level*English not well/not at all			1.130 (7.53)			0.692 (11.04)
Occupation dummies	N_Inc.	Inc.	N_Inc.	N_Inc.	Inc.	N_Inc.
Adjusted R^2	0.352	0.374	0.352	0.389	0.420	0.393
Sample size	532,109	532,109	532,109	84,172	84,172	84,172

Source: 2000 US Census, 1% Public Use Microdata Sample (PUMS)

N_Inc. variable not included, *Inc.* variable included

[a]Level of English required in the occupation of employment according to the O*NET database, scaled by 100

Table 2 OLS and IV estimates of partial effect on earnings of occupational English-language requirements by nativity, 2000 US Census

Method of estimation	Native born		Foreign born	
	Without occupation fixed effects	With occupation fixed effects	Without occupation fixed effects	With occupation fixed effects
IV	0.031 (61.24)	0.067 (41.47)	0.041 (59.78)	0.100 (27.57)
OLS	0.013 (112.51)	0.013 (67.88)	0.016 (56.54)	0.011 (26.40)
Sample size	532,109	532,109	84,172	84,172

Scaled level of English required in the occupation of employment according to the O*NET database. Source: 2000 US Census, 1% Public Use Microdata Sample (PUMS)

discussed above are reviewed: the inclusion of occupation fixed effects and the interaction terms between the individual's proficiency in English ($PROF_i$) and the English requirements of his job (ER_i). These results are presented in Table 1. Findings from the IV analysis and for the Dickens and Katz (1987) model are presented in Tables 2 and 4, respectively. Table 3 lists results from the intermediate step (estimation of Eq. 3a) in the (Dickens and Katz 1987) model. Comment on findings from models that allow for nonlinearities in the relationship between earnings and the English-language requirements of occupations is provided in a footnote.

Results from the benchmark model of Eq. 1 are listed in column (i) of Table 1. The coefficients on the standardizing variables in the vector X are reasonably close to the evidence reported in recent studies. Thus, the payoff to one additional year of education for the native born is about 7%, and the payoff for the foreign born is only one half of this (i.e., about 3.5%).[6] The relative magnitudes of these payoffs for the native born and foreign born are consistent with evidence reported in Chiswick and Miller (2008). Reflecting the use of information on occupational English requirements in the current study, the payoffs to education in Table 1 are about 30% less than those in the comparison study.

The continuous variable that records the natural logarithm of weeks worked in 1999 indicates an elasticity of earnings with respect to weeks worked close to unity for the native born and around 0.87 for the foreign born. The marital status variable shows that married men earn more than their nonmarried counterparts, with their earnings advantage being 26% for the native born and 20% for the foreign born. These weeks worked and marital status effects on

[6]The payoff to potential labor market experience is 2.3% for the native born and 1.0% for the foreign born, when evaluated at 10 years. Similar results are documented by Chiswick and Miller (2008), where the payoff to potential labor market experience in the absence of control for the occupational English requirements was reported to be 2.3% for the native born and 0.8% for the foreign born.

Table 3 Earnings functions with and without occupation fixed effects, by nativity, 2000 US Census

Variable	Native born		Foreign born	
	Without occupation fixed effects	With occupation fixed effects	Without occupation fixed effects	With occupation fixed effects
Constant	4.346 (192.77)		5.850 (108.41)	
Log weeks worked	1.010 (183.25)	0.957 (370.13)	0.875 (73.01)	0.842 (152.71)
Years of education	0.106 (201.74)	0.058 (103.06)	0.053 (69.56)	0.023 (29.05)
Experience	0.033 (71.47)	0.032 (75.94)	0.011 (10.99)	0.019 (19.00)
Experience squared/100	−0.056 (55.29)	−0.055 (63.35)	−0.016 (8.06)	−0.030 (16.31)
Married	0.270 (111.26)	0.203 (87.94)	0.214 (35.81)	0.159 (28.31)
South	−0.057 (24.71)	−0.075 (34.39)	−0.072 (11.75)	−0.089 (15.44)
Metropolitan area	0.211 (35.57)	0.163 (30.63)	0.135 (5.02)	0.085 (3.25)
Veteran	−0.045 (16.52)	−0.045 (17.52)	−0.083 (6.28)	−0.046 (3.69)
Blacks	−0.153 (42.85)	−0.076 (21.28)	−0.185 (17.21)	−0.067 (6.53)
Years since migration (YSM)			0.011 (13.14)	0.012 (15.69)
YSM squared/100			−0.011 (5.66)	−0.014 (8.18)
English very well	−0.059 (11.56)	−0.040 (8.26)	−0.080 (8.53)	−0.057 (6.95)
English well	−0.097 (7.46)	−0.070 (5.93)	−0.261 (26.20)	−0.134 (15.05)
English not well/not at all	0.013 (0.70)	0.010 (0.59)	−0.374 (34.99)	−0.224 (23.31)
Adjusted R^2	0.335	0.412	0.363	0.449
Sample size	532,109	532,109	84,172	84,172

Each occupation is assigned an intercept term. Source: 2000 US Census, 1% Public Use Microdata Sample (PUMS)

Table 4 Effects of occupational English requirements on occupation fixed effects

	Native born	Foreign born
Constant	4.792 (103.57)	5.819 (104.32)
English requirements/100	0.954 (9.97)	1.024 (8.29)
$\overline{R}^2$	0.171	0.148
Sample size	505	487

Occupation fixed effects are from regressions in Table 3 columns (ii) and (iv). Source: 2000 US Census, 1% Public Use Microdata Sample (PUMS)

earnings are virtually identical to those reported where occupational English requirements are not held constant.

The estimated partial effects show that native-born men who speak a language other than English at home and who speak English very well or well have earnings 7% to 11% lower than those who only speak English at home (see Fry and Lowell 2003 and Chiswick and Miller 2007 for related and consistent evidence). Among foreign-born men, those who speak a language other than English at home and who speak English very well earn 7% less than those who speak only English at home, while those who speak English well have about 20% lower earnings, and those who speak English not well or not at all have earnings 30% less than the reference group that speaks only English at home. These partial effects are greater in absolute value than those found using estimating equations that do not control for the occupational English requirements: for the foreign born they are 20–30% greater, and for the native born, they are a more modest 10% greater than the absolute value of the effects reported in Chiswick and Miller (2008) for a similar specification of the earnings equation (but without the ER_i variable). The fact that the Table 1 column (i) estimates are greater in absolute value than the results in Chiswick and Miller (2008) points to a positive correlation between being proficient in English and working in occupations that have high required levels of English. The fact that the difference in this regard is reasonably modest for the native born and only 20–30% for the foreign born indicates that the impacts on earnings of English-language proficiency are associated primarily with intra-occupational effects.

The English-language requirements of the occupation are positively associated with earnings. Thus, the coefficient on the O*NET level of English variable is 0.013 for the native born and 0.016 for the foreign born. These impacts are sizeable. A one-standard-deviation change in the required level of English would be associated with 15.9% higher earnings for the native born and 19.7% higher earnings for the foreign born. These changes are the equivalent of the earnings effects associated with 2 years of schooling for the native born and almost 6 years of schooling for the foreign born. A change from an occupation with a required level of English of 20 (e.g. Metal refining furnace operators and tenders) to one with a required level of English of 80 (e.g., Postsecondary environmental science teachers; where 20 and 80 are the practical extremes of the data) would be associated with a 78 percentage point

change in earnings for the native born and with a 96 percentage point change in earnings for the foreign born, other measured variables being the same.[7]

Column (ii) of Table 1 lists results from the model that includes occupation fixed effects measured at the level of the broad Census occupational category. This augmentation of the estimating equation is associated with an increase of two percentage points in the explanatory power of the model for the native born and with an increase of three percentage points in the explanatory power of the model for the foreign born. The estimated coefficients on the occupation variables (not reported here) are sizeable, revealing a difference in *ceteris paribus* earnings across occupations of 70 percentage points for the native born, and of 83 percentage points for the foreign born. However, this augmentation of the earnings equation has little impact on the estimate of the relationship between earnings and the occupational English-language requirements for the native born. It is associated with a 30% reduction in the impact of the English requirements variable for the foreign born. In this augmented specification, the partial effects on earnings of the scaled occupational English-language requirements variable is 1.337 for the native born and 1.107 for the foreign born. In other words, the substantial effects of the English-language requirements of the occupation on earnings are not simply capturing the standard inter-occupational wage structure.

The effects of the English-language requirements of the occupation in column (i) of Table 1 are only about one half of the effects found in a simple regression of earnings on the ER_i variable. This implies a strong association between the ER_i variable and the other regressors in the multivariate model, a pattern that can be explored using the standard omitted variables formula (see Johnson and Solon 1986 for a related application). Educational attainment was the chief contributor to the diminution of the effects of the ER_i variable between the simple and multivariate regression, accounting for about two thirds of the change for the native born and around 57% of the change for the foreign born. The weeks worked variable accounted for a further 20% of the change. English proficiency variables accounted for around 25% of the diminution for the foreign born but had little impact for the native born. The findings for educational attainment and for English proficiency among the foreign born indicate a positive sorting of better skilled workers into occupations that have more demanding English-language requirements. This would indicate a matching of worker and job attributes. This matching is presumably also reflected in the relative rates of remuneration that workers with different levels of English can obtain in occupations requiring low and high levels of English-language skills.

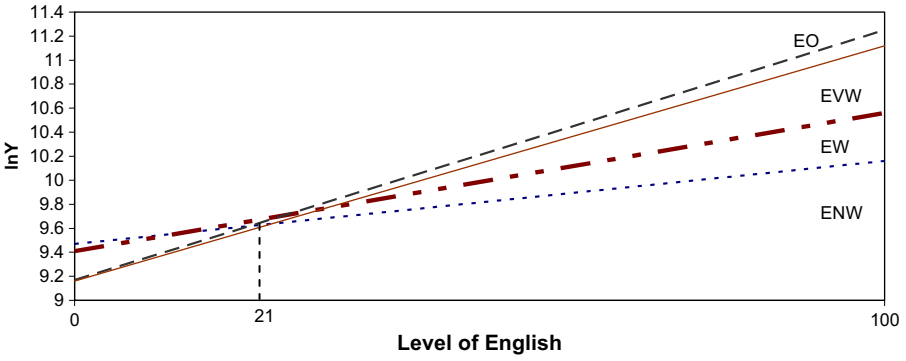
[7] Including the ER_i variable in the estimating equation in quadratic form does not alter the findings: the results show that earnings increase at a decreasing rate with the English requirements of jobs for the native born: $1.964ER_i - 0.690ER_i^2$ and at an increasing rate for the foreign born: $0.935ER_i + 0.649ER_i^2$. All coefficients were statistically significant. The other estimates in the model were largely unaffected by the inclusion of the squared ER_i variable.

To examine this issue, interaction terms between the variables for O*NET English requirements and the workers' self-reported proficiency in English were included in the estimating equation. These results are reported in column (iii) of Table 1. While the same sets of analyses are undertaken for the native born and for the foreign born, the discussion will concentrate on the findings for the foreign born because nearly all of the native born speak only English.

The earnings-English requirements profiles for the four levels of English proficiency are presented in Fig. 3.

Figure 3 makes it very clear that the earnings gains associated with the English requirements of the job depend on the individuals' proficiency in English. For English-only speakers, there is a gain of 2.1 log points in earnings across the level of English scale and a gain of 25.5% associated with a one-standard-deviation change in the English requirements variable. The gain for those who speak a language other than English at home and who speak English very well is slightly smaller, 2.0 log points across the 0–100 scale and 24.1% for a one-standard-deviation change in the ER_i variable. These effects are not significantly different. However, the increases in earnings across the O*NET level of English scale for those who speak English "Well" is only 1.1 log points (and 14.1% for a one-standard-deviation change)—about one half the gains for English-only speakers. Finally, the increases in earnings across the O*NET level of English scale for those who speak English "Not Well or Not at All" is even smaller, being only 0.7 log points (8.5% gain for a one-standard-deviation change), about one third of the gains for English-only speakers.

Those who speak a language other than English at home and who have more limited English skills (i.e., they speak English Well, Not Well, or Not



EO = English Only;
EVW = English Very Well;
EW = English Well;
ENW = English Not Well/Not at All.

Source: Authors' calculations based on Table 1.

Fig. 3 Predicted earnings by level of English for English proficiency groups, foreign born

at All) actually have higher earnings than the benchmark group of English-only speakers if they are in jobs with minimal English requirements (e.g., Floor Sanders and Finishers, with a score of 9). However, once they are in jobs that require a level of English of 21 or more, those with limited English skills are at an earnings disadvantage (e.g., Janitors and Cleaners have a score of 23, Postal Service Mail Sorters have a score of 29). This disadvantage rises rapidly across the scale for required levels of English. This suggests a negative selectivity on other characteristics (perhaps unmeasured ability) of English-only speakers in jobs that require no English-language proficiency.

The rewards to correct matching of the skills of workers and the requirements of jobs are very apparent in the US labor market. The pattern also characterizes analyses for the native born, although these differences are compressed, perhaps because only a very small percentage of the native born speak a language other than English.

Thus, there is an interaction effect on earnings between the respondent's proficiency in English and the level of English required for the occupation. Those with higher levels of English-language proficiency earn more, but the increase in earnings is greater if they are in an occupation that requires greater proficiency in English. Those who are proficient in English but in occupations that do not require English-language skills are not taking advantage of a skill that they have.

3.2 Estimation

An oft-repeated concern in analyses of earnings is that many of the standard regressors are potentially endogenous. Included are educational attainment, marital status, occupational choice, and the English requirements variable. Addressing these concerns using IV estimation is typically not practical, as few data sets contain suitable instruments. In Chiswick and Miller's (1992) study of the returns to English-speaking proficiency, instrumenting the English proficiency variable (with veteran status, number of children, and the concentration of people with similar mother tongues as instruments) generated estimates which were considered unusually large. For example, the partial effect on earnings of proficiency in English changed from 0.169 to 0.571 for analysis of 1980 US Census data. Moreover, the choice of instruments has been criticized on the grounds that they are likely to be correlated with the immigrant's earnings capacity (see, for example, Borjas 1994, p.1684).

In the present study, the variable of most interest is the ER_i variable. While the data set is, like previous studies, limited in the choice of instruments, the sensitivity of the findings to the endogeneity issue was explored using the approach outlined in Section 2. Selected findings are presented in Table 2 together with comparison ordinary least squares (OLS) results.

The Table 2 results have two features, and these follow findings reported in the related literature (e.g., Chiswick and Miller 1992, 1995). First, the IV estimates of the partial effect of occupational English-language requirements on earnings are positive and exceed the OLS estimates. Second, the difference

between the IV and OLS estimates is uncomfortably large, ranging from a ratio of 2.4 to one of 9.1. Thus, this exploration suggests that the already sizeable OLS estimate of the effect on earnings of occupational English-language requirements is likely to be lower bound. It also suggest that the instruments available for the analysis is an important caveat to the findings.

There is a further finding of note from the IV analysis, and this is in relation to the estimates of Eq. 2b, which is the first stage of the IV analysis. This showed that educational attainment is the primary determinant of being in an occupation that has a higher English-language requirement. There was a difference of 30 points between the English-language level requirements of occupations of the native born with the lowest and highest educational attainments and yet a difference in the level of the English requirements variable of less than one point among English-only native-born speakers and those who had limited English skills. Among the foreign born, the range in the estimates was 25 points for educational attainment and four points for English speaking proficiency. These findings are, not surprisingly, consistent with the results discussed above in relation to the Johnson and Solon (1986) decomposition.[8]

3.3 Dickens and Katz 1987 robustness analysis

The first step in implementation of the Dickens and Katz (1987) model is to estimate Eq. 3a. This differs from the specification of column (ii) of Table 1 in two ways: (1) the occupational fixed effects are measured at a much finer disaggregation (about 500 occupational categories); and (2) the occupational English-language requirements is not included in the model. Table 3 lists the results of the earnings equations without the occupational English requirements but both with and without the detailed dichotomous occupation (fixed effects) variables.

The incorporation of occupation fixed effect variables into the regression equation (504 dichotomous variables for the native born and 491 for the foreign born)[9] raises the adjusted R^2 by eight percentage points for the native born (a 23% increase in the explanatory power of the equation) and by nine percentage points for the foreign born (a 24% increase in the explanatory power of the equation). This is a much greater improvement in the explanatory power than was associated with the inclusion of dummy variables for the broad (23) Census occupational categories in Table 1. The change in specification is generally associated with a reduction (in absolute value) of the estimated impacts on earnings of the individual-level variables. For example, the esti-

[8]Related to this, the partial effect of education on earnings in the IV estimation (not reported in Table 2) is very small: 2% for the native born and 1% for the foreign born. This is further reason for focus on the OLS results.

[9]While the Census occupational categorization has 509 codes, there are several occupations that are not represented in the sample used. Included in this group are the military occupations for which the O*NET database does not collect information.

mated effect of years of education on earnings falls by about one half once the occupational fixed effects are included in the model. This is a consequence of variables that are associated with higher earnings, such as education, also being associated with above average representation in high-paying occupations.

There is considerable variation in the magnitude of the coefficients of the occupation dummy variables in the augmented earnings equation. Table 4 explores the links between the coefficients of the occupation fixed effects on earnings and the occupational English requirements. It shows that English requirements account for between 15% and 17% of the variation in the occupation fixed effects.

In each earnings regression, the English requirements variable is highly significant. However, reflecting the lower degrees of freedom when the data are analyzed at the higher level of aggregation, the t statistics are far smaller than those that were reported in Table 1. Nevertheless, with the smallest t being 8.29, the significance of the English requirements variables is not in doubt.

The estimated impacts on earnings of the information on the English requirements, as per the analysis of the occupation fixed effects in Table 4, are sizeable, though about 0.3 to 0.5 of a percentage point less than the effects estimated on the basis of mixing aggregate-level data with the micro-level data (see Table 1). This suggests merit to the robustness checks reported here. But it also shows that the fundamental theme of the study, that English-language requirements are very important to the understanding of variations in earnings, carries across the alternative sets of analyses reported here.

4 Conclusion

This paper examines the required English-language proficiency of occupations in the US labor market. It also estimates the value of proficiency in English by evaluating its impact on earnings for men aged 25 to 64 years who worked in 1999. It extends previous research by its use of measures of English-language requirements in the O*NET database and its exploration of the interactions among these measures, worker characteristics, and earnings.

The O*NET database provides information on the standardized scores of the importance and level of English-language skills in each of over 800 occupations. The importance and level scores are very highly correlated, both overall ($r = 0.92$) and within each of the six broad occupational groups considered. These occupational English scores were linked to data from the 2000 US Census, and regression analyses of earnings are conducted separately for the native born and the foreign born.

The results show that earnings increase with the English-language requirements of the occupation. A one-standard-deviation change in the required level of English is associated with 15.9% higher earnings for the native born and 19.7% higher earnings for the foreign born. These changes are the equivalent of the earnings effects associated with 2 years of schooling for the

native born and almost 6 years of schooling for the foreign born. Inclusion of the information on the English-language requirements of the occupation has no impact on the links between the individual's proficiency in English and earnings among the native born. It is, however, associated with a diminution, by 20–30%, in the partial effect of English proficiency on earnings among the foreign born. In other words, proficiency in English is associated more with intra-occupational increases in earnings than it is with inter-occupation increases in earnings.

The results reported are robust with respect to the alternative specifications of the estimating equations (with occupation fixed effects, with the English-language requirements entered in quadratic form) and the alternative methods of estimation (IV, Dickens and Katz 1987 procedure) explored.

Interaction terms between the self-reported English-language proficiency of workers and the English-language requirements of the occupations in which they work reveal that there are rewards to correct matching of worker language skills and job requirements in the US labor market. Workers with poor English skills do relatively well when employed in jobs that have very low English-language requirements. They do relatively poorly when employed in jobs that have high English-language requirements. These results are very striking for the foreign born but also characterize the determinants of earnings for the native born.

For both birthplace groups, labor markets appear to sort workers appropriately, with those with high levels of English proficiency tending to work in jobs which require a high level of proficiency and in which English-language proficiency is important.

Thus, there is a complementarity in occupational choice (i.e., English-proficient workers tend to work in jobs requiring proficiency) and in generating earnings (i.e., English-proficient workers earn more than those not proficient, and this effect is greater the higher the level of proficiency required by the occupation).

Acknowledgements We thank Derby Voon for research assistance and three anonymous referees for helpful comments. Chiswick acknowledges research support from the Smith Richardson Foundation and the Institute of Government and Public Affairs, University of Illinois. Miller acknowledges financial assistance from the Australian Research Council.

Appendix A Description of variables

The variables used in the statistical analyses are defined below. Mnemonic names are also listed where relevant.

Data source

2000 US Census of Population, 1% Public Use Microdata Sample; O*NET Occupational data from the O*NET Consortium described in footnote 1.

Definition of population

Native-born and Foreign-born men aged 25 to 64 in the 50 States and the District of Columbia.

Dependent variable

Earnings

The natural logarithm of the individual's annual earnings from wage and salaried employment or self-employment for 1999.

Explanatory variables

Educational attainment (EDUC)

The total years of full-time equivalent education. The following values are assigned to the Census categories: completed less than fifth grade (2 years); completed fifth or sixth grade (5.5); completed seventh or eighth grade (7.5); completed ninth grade (9); completed tenth grade (10); completed 11th grade (11); completed 12th grade or high school (12); attended college for less than 1 year (12.5); attended college for more than 1 year or completed college (14); Bachelor's degree (16); Master's degree (17.5); Professional degree (18.5); Doctorate (20).

Labor market experience (EXP)

A measure of potential labor market experience, computed as age - years of education - 6.

Log weeks worked (WEEKS)

The natural logarithm of the number of weeks the individual worked in 1999.

Years since migration (YSM)

This is computed from 2000 minus the year the foreign-born person came to the USA to stay.

English-language fluency

Three dichotomous variables that distinguish individuals who speak a language other than English in the home and who speak English either: (1) "Very Well"; (2) "Well"; (3) "Not Well"/"Not at All". The benchmark group is those who speak only English at home.

Race (BLACK)

This is a dichotomous variable, set to one if the individual is black and set to zero for all other racial groups.

Marital status (MARRIED)

A dichotomous variable that distinguishes individuals who are married, spouse present (equal to 1) from all other marital states.

Location

The two location variables record residence in a non-metropolitan area (NONMET) or in a Southern State (SOUTH). The states included in the latter are: Alabama, Arkansas, Delaware, District of Columbia, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, Missouri, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, West Virginia.

Veteran (VETERAN)

This is a dichotomous variable, set to one if the individual is veteran of the US Armed Forces, and set to zero otherwise. Those currently in the military are excluded from the sample.

English requirements (LEVEL and IMPORTANCE)

These variables record the scores for the level and importance of English requirements for each occupation obtained from the O*NET database. When the O*NET database was first developed, job analysts relied on information from the *Dictionary of Occupational Titles* and modified this to suit the set of occupational codes (Standard Occupational Classification System) used in O*NET. Since June 2001, data have been collected from workers in targeted subsets of the occupations identified in O*NET, using a two-stage sampling design based on random samples of workers in targeted occupations within a random sample of businesses. These survey data have been progressively integrated into the initial O*NET database. This study used version 12.0 released in September 2007. Nearly all of the occupational language requirements are based on worker assessments for their occupation. There are no data in O*NET on military occupations.

References

Borjas G (1994) The economics of immigration. J Econ Lit 32(4):1667–1717

Chiswick BR, Miller PW (1992) The immigrant experience in Canada and the United States. In: Chiswick BR (ed) Immigration, language and ethnic issues: Canada and the United States. American Enterprise Institute, Washington DC, pp 229–296

Chiswick BR, Miller PW (1995) The endogeneity between language and earnings: international analyses. J Labor Econ 13(2):246–288
Chiswick BR, Miller PW (2007) The economic cost to native-born Americans of limited English language proficiency. In: Chiswick BR, Miller PW (eds) The economics of language: international analyses. Routledge, London, pp 413–430
Chiswick BR, Miller PW (2008) Why is the payoff to schooling smaller for immigrants? Labour Economics (in press)
Dickens WT, Katz LF (1987) Inter-industry wage differences and industry characteristics. In: Lang K, Leonard JS (eds) Unemployment and the structure of labor markets. Basil Blackwell, New York USA, pp 48–89
Duncan OD (1961) A socioeconomic index for all occupations. In: Reiss AJ Jr (ed) Occupational and social status. Free Press, New York, pp 109–138
Fry R, Lowell BL (2003) The value of bilingualism in the US labor market. Ind Labor Relat Rev 57(1):128–140
Groshen EL (1991) Sources of intra-industry wage dispersion: how much do employers matter? Q J Econ 106(3):869–884
Gyourko J, Tracy J (1988) An analysis of public- and private-sector wages allowing for engogenous choices of both government and union status. J Labor Econ 6(22):229–253
Hartog J, Oosterbeek H (1993) Public and private sector wages in the Netherlands. Eur Econ Rev 37(1):97–114
Heckman JJ, Lochner LJ, Todd PE (2003) Fifty years of mincer earnings regressions. NBER Working Paper No. 9732
Johnson G, Solon G (1986) Estimates of the direct effects of comparable worth policy. Am Econ Rev 76(5):1117–1125
Meng X, Gregory RG (2005) Intermarriage and the economic assimilation of immigrants. J Labor Econ 23(1):135–175
Moulton BR (1986) Random group effects and the precision of regression estimates. J Econom 32(3):385–397