Preliminary Draft
Not for quotation nor citation **April 2011**
without an author's permission

MATCHING LANGUAGE PROFICIENCY TO OCCUPATION: THE EFFECT ON IMMIGRANTS' EARNINGS *

Barry R. Chiswick
Department of Economics
George Washington University
and
IZA-Institute for the Study of Labor

and

Paul W. Miller
School of Economics and Finance
Curtin University

Keyword: English Language, Occupation, Earnings, Immigrants, Schooling

JEL Codes: J24, J31, F22

* We thank Derby Voon for research assistance and William P. Bridges for helpful comments. Chiswick acknowledges research support from the Smith Richardson Foundation. Miller acknowledges financial assistance from the Australian Research Council.

February 2011

MATCHING LANGUAGE PROFICIENCY TO OCCUPATION: THE EFFECT ON IMMIGRANTS' EARNINGS

ABSTRACT

This paper analyzes the effect on earnings of the matching of English language skills to occupational requirements or occupational norms for adult male immigrants. It uses data from the Occupational Information Network (O*NET) database and a "Realized Matches" procedure to quantify expected levels of English skills in each of over 500 occupations in the US Census. Earnings data from the 2000 US Census for foreign-born adult male workers are then examined in relation to these occupational English requirements or norms using the Over/Required/Under (or ORU) technique developed for the study of schooling. The analyses show that earnings are related to a "correct" matching of an individual's language skills with what is expected in his occupation. Mismatches have a small effect on earnings – positive for extra proficiency and negative for deficits in proficiency, relative to the norm in the occupation. The findings are robust with respect to a range of measurement and specification issues.

February 2011

MATCHING LANGUAGE PROFICIENCY TO OCCUPATION: THE EFFECT ON IMMIGRANTS' EARNINGS

I. INTRODUCTION

One skill that appears to be very important in the US labor market, particularly among immigrants, is knowledge of the English language. Chiswick and Miller (1995), for example, show that immigrants in the US who are proficient in English have earnings about 17 percent higher than immigrants with limited English skills, other measured variables the same. This is the equivalent of about three years of schooling among immigrants. Similar patterns have been found for other immigrant receiving countries (Chiswick and Miller, 2007, and the references therein). To date, however, the research in this area has not examined whether the earnings premium to language proficiency varies according to the language requirements or norms in the jobs (occupations) in which immigrants work.[1]

This paper addresses, for the United States, the issue of the extent to which the English language requirements in the respondent's occupation influence the respondent's earnings. Moreover, it is concerned with the effects on earnings of the discrepancy between the respondent's proficiency and the requirements in his occupation. The data under study are for adult men, age 25-64, from the US 2000 Census. In these data there is variation in the occupational English language requirements, and there is variation in the respondent's English proficiency for the foreign born. Among the native-born adults, however, there is virtually no variation in the respondent's proficiency; nearly all report that they are monolingual English language speakers, or if they speak another language at home, nearly all report they speak English "very well." Hence, the analysis of the discrepancy between the English language proficiency of the respondents and the requirements of their occupations is limited to the foreign born.

[1] Chiswick and Miller (1998a) provide some information on this in the context of a study of the earnings of native born bilingual workers in the US.

It is hypothesized that earnings increase with the level of English language skills that are required or the norm in the person's occupation, other measured variables the same, for both the foreign born and the native born. It is also hypothesized that among the foreign born the proficiency of an individual greater than this level has a smaller positive effect on earnings, while proficiency levels below this level have a negative effect on earnings that is smaller in absolute value than that for the required/norm level. The occupational level is taken as given for the purpose of this analysis. The respondents' English language proficiency is also taken as exogenous. [2, 3]

The methodology used to analyze the discrepancy between the person's proficiency and that required in his occupation is adopted from the education literature (Hartog 2000, Duncan and Hoffman 1981, Rumberger 1987, Chiswick and Miller 2009). Whereas the ORU (overeducated/required education/undereducated) literature refers to the individual's years of schooling and the requirements in his occupation (usually mean or mode), in this analysis the respondent's self-reported proficiency in English is compared to the English language requirements or norms in his occupation. Just as the analysis for education has shown that earnings vary systematically with the occupational required level and discrepancy (mismatch) in education, so too it is shown here that earnings vary systematically with the occupational level of required proficiency in English and the individual's deviation from this proficiency.

Section II is a discussion of the language requirements or norms of occupations. Ways of relating information on an individual's self-reported language proficiency to the language requirement of his or her occupation are assessed in Section III, with the main

[2] There are undoubtedly unmeasured variables that account for why there are educational and linguistic mismatches, that is, why some individuals appear to be overqualified (underqualified) given their measured skills compared to others in their occupation. Differences in cohorts (younger versus older workers), unmeasured dimensions of ability or quality of skills, random events, and among the foreign born, the international transferability of skills, are presumably relevant. This is the subject of ongoing research, but is beyond the scope of this paper.

[3] For analyses of the determinants of destination language proficiency among immigrants, see Chiswick and Miller (2007).

issues being illustrated using data from the 2000 US Census, one percent microdata file. Section IV presents an analysis of earnings, with the focus on the matching and discrepancy of language skills to occupation. A number of robustness checks are considered in Section V. Section VI offers a summary and conclusion.

II. OCCUPATIONAL ENGLISH LANGUAGE REQUIREMENTS[4]

The O*NET Database

The Occupational Information Network, or O*NET, database, records an extremely wide range of characteristics for nearly all narrowly defined occupations, including the level of English required in a wide range of occupations.[5] In particular, the O*NET database contains details about the "Knowledge of the structure and content of the English language, including the meaning and spelling of words, rules of composition, and grammar". Two sets of information were collected. The first is about "How important is knowledge of the ENGLISH LANGUAGE to the performance of *your current job*?" (emphasis in original). 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 the English language needed to perform the current job was collected only among those who felt that English was somewhat or more important to performance in this 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

[4] Parts of this presentation are from Chiswick and Miller (2010), which provides the first formal assessment of the links between earnings and occupational language requirements using the O*NET data. This earlier study did not, however, adopt the ORU perspective.

[5] The National O*NET Consortium was organized to develop the Occupational Information Network (O*NET) and its related products for the Employment and Training Administration of the US Department of Labor.

current job were coded as zero on the index for level. Hence an eight-point scale (0-7) results.

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

$S = ((O - L) / (H - L)) * 100$

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 original scores on the five-point scale of importance become 1 = 0; 2 = 25; 3 = 50; 4 = 75; and 5 = 100. The scores on the eight-point scale of English level become 0 = 0; 1 = 14.3; 2 = 28.6; 3 = 42.9; 4 = 57.1; 5 = 71.4; 6 = 85.7; 7 = 100.[6]

When the O*NET database was first established, job analysts used 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. From June 2001, data have been collected from workers in targeted subsets of the occupations identified in O*NET, using a two-stage sampling design (random samples of workers in targeted occupations within a random sample of establishments). These survey data have been progressively integrated into the initial O*NET database. In the September 2007 release, O*NET Version 12.0, used in this study, virtually all occupations had data based on surveys of each occupation's incumbents. Hence, the job requirements obtained from the O*NET database should be viewed as having been compiled using the Worker Self-Assessment approach identified in the ORU literature.

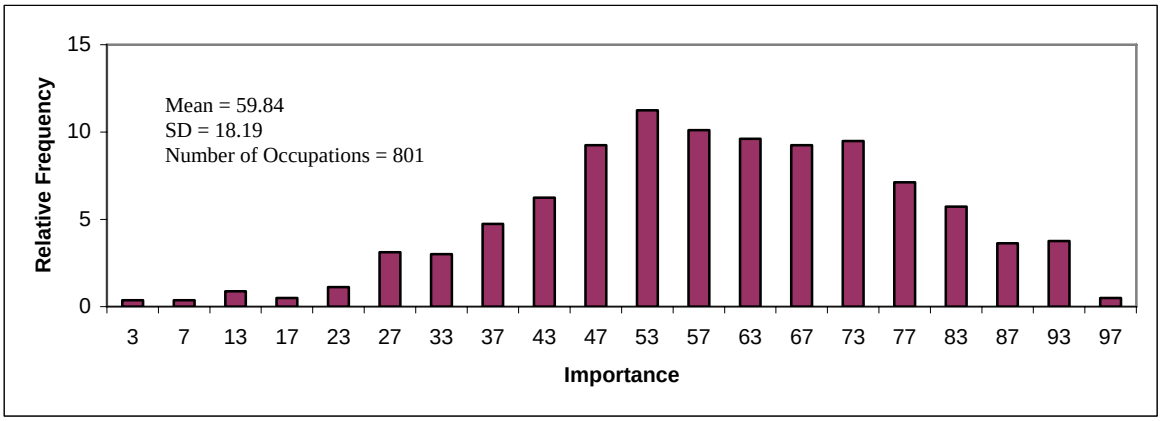
[6] Occupations with a level of English around 20 include glaziers and crossing guards. Those with a level of English around 80 include postsecondary environmental science teachers and sociologists. Occupations with an importance of English around 20 include logging equipment operators and models, while those with an importance of English around 80 include first line supervisors/managers of correctional officers and respiratory therapists. Economists have a level of English of 73 and an importance of English of 91, while for sociologists the scores are 78 for level and 84 for importance.

Importance of English

There is considerable variation in the importance of knowledge of the English language to job performance (Figure 1). The mean standardized score is 59.84, which is almost half-way between the "Important" and "Very important" points of the scale used in data collection. The standard deviation is 18.19, which is the equivalent of a change of almost one category on the underlying five-point scale. Moreover, the importance varies from minimal amounts (standardized scores around ten) in some occupations, to occupations where knowledge of English is very important. Occupations where English is not important include "Paperhanger" (Score of 8), "Precious Metal Workers" (13) and "Continuous Mining Machine Operator" (13). Examples of occupations where English is very important are "Public Relation Managers" (96), "Proof Readers and Copy Makers" (95) and "Judges, Magistrate Judges and Magistrates" (95). Weighted means of scores are used to go from the very detailed occupations to these broader categories.

The relative frequency distribution in Figure 1 shows that the occupations in the US labor market cover a full range of values on the standardized measure of the importance of the English language.

Figure 1
Relative Frequency of Occupations on Standardized Score of Importance of English

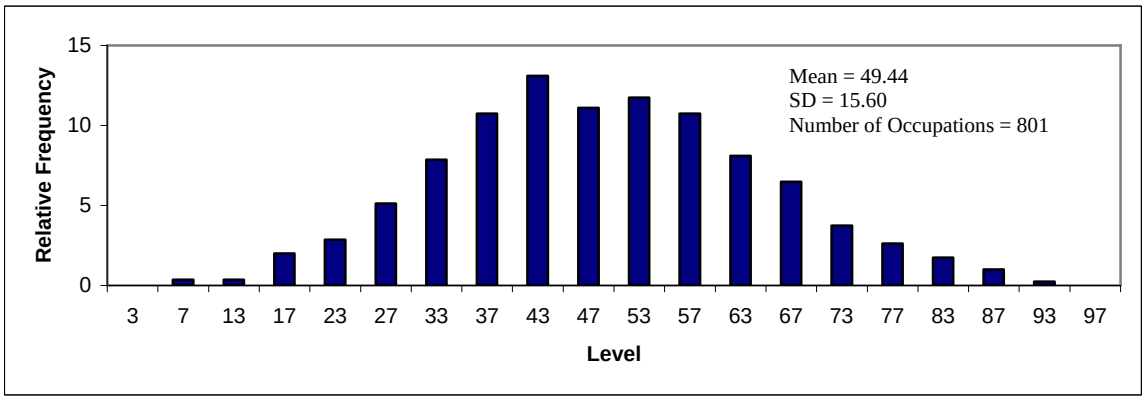


Note: The 801 occupations for which details are available in the O*NET database are used in the compilation of the data for Figures 1 and 2. They were not weighted by the number of workers in the occupation, and hence the distribution is sensitive to how jobs are grouped into occupations. The standardized scores have been collapsed into bands of width five (*e.g.*, 46-50, 51-55, 56-60). A mid-point is used on the horizontal axis.
Source: O*NET Database.

Level of English

The data on the level of English required for all occupations are illustrated in Figure 2. The mean for all occupations is 49.44, which is two-thirds of the way between levels 2 and 4 (on the original scale) 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 15.60. The occupations in the US labor market cover a wide range of the standardized scores, although there is very limited representation above scores of 80 and below scores of 20. Compared to the frequency distribution for the importance dimension, the data for the level of English needed to perform the job tend to be bunched more in the bottom half of the standardized scores. Nevertheless, there is a very high correlation across the 801 occupations between the scores for the importance of English and the level of English: it is 0.92 (not weighted by the sample size in the occupation). 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.

Figure 2
Relative Frequency of Occupations on Standardized Score of Level of English



See Notes to Figure 1.
Source: O*NET Database.

Robustness Check

The empirical analysis in Section IV is based primarily on the reference levels of English computed from the O*NET database described above. The robustness of the

empirical findings is examined using an alternative measure for the English language requirements of each occupation, in particular, the mean English proficiency of incumbents in the occupations. This is a Realized Matches way of benchmarking the skill requirements of occupations. The mechanics of this will be apparent from the discussion in Section III. In addition, consideration is given to scaling the data on the O*NET English requirements and workers' actual English proficiency so that they have the same mean and standard deviation. Comparison of variables with similar distributions has appeal from the perspective of measurement theory. These measurement issues are discussed in greater detail in Section V.

III. LINKING WORKER ATTRIBUTES TO JOB REQUIREMENTS

The standardized O*NET scores on the "Level of English" represent the general requirements for this skill for occupations in the US labor market.[7] They provide a reference point for assessing whether workers have, relative to others, "too much" English for the job, "too little" English, or the right amount of English language proficiency.

Information on the English language skills of workers is collected in many censuses and surveys, and these data use a variety of response categories (see Chiswick and Miller, 1998b). In this study data from the 2000 US Census are used. The 2000 US Census permits self-reported English proficiency to be categorized using a five-interval scale. The highest level of this scale is 5 = Speaks only English at home. All other levels relate to individuals who speak a language other than English at home, and self-report speaking English: 4 = Very Well; 3 = Well; 2 = Not Well; 1 = Not at All. The analysis which follows requires the Census English proficiency data to be converted to a continuous scale, between 0 and 100, the same as that used in the O*NET database. Three alternative scorings are considered in this analysis.

[7] These can be thought of as national averages, recognizing that there can be regional and ethnic enclave differences, and even differences across firms within a region.

The first of these follows Evans (1986)(1987), who uses the scores: 0 = Speaks no English at all; 33 = Speaks English "Not Well"; 67 = Speaks English "Well"; 100 = Speaks English "Very Well" or speaks only English at home. This is based on Evans's findings that the effects of English proficiency variables on occupational attainment were approximately linear. The grouping for the score of 100 has support in the literature on the ranking of these categories (see, for example, Kominski, 1989). Espenshade and Fu (1997, p.293) argue that "…there is not much difference in English proficiency between immigrants who use a language other than English at home but who say they speak English "Very Well" and those who use only English at home". They therefore group these two categories together in their statistical analyses. Bleakley and Chin (2004) also follow this categorization.

The second scoring separates the English "Very Well" and "English only group" and treats the English only group as having the higher level of proficiency. This has a motivation similar to the discussion in Espenshade and Fu (1997), that one might create a higher category for the English only speakers, such as English "Extremely Well". Chiswick and Miller (2008a), for example, distinguish those who speak English only and second language speakers who speak English very well. Their statistical results suggest there are some, albeit apparently relatively modest, advantages to this disaggregated approach. Within this five-category setting, values of 0, 25, 50, 75 and 100 are assigned to the English proficiency levels, with these values reflecting an extension of the linear scoring proposed by Evans (1986, 1987).

Finally, the differences in the mean logarithmic earnings of immigrants in the English proficiency categories were examined, and these differences were used to establish an alternative weighting scheme. This weighting was surprisingly similar to that advocated by Evans (1986, 1987), with the values being 0, 27, 62, 97 and 100. Table 1 provides a summary of the alternative values assigned to the Census data on English proficiency.

Table 1
Relative Values (Weights) Assigned to Census Information on Proficiency in English

Proficiency in English	Evans'(1986)(1987) Proficiency Values (i)	Alternative Values (ii)	Values Based on Mean Earnings (iii)
English Only	100	100	100
Very Well	100	75	97
Well	67	50	62
Not Well	33	25	27
Not at All	0	0	0

Regardless of the scale used for the respondent's actual proficiency, the actual and "required" or reference proficiency are separate scales, each of which ranges from zero to 100. The difference in these scales refers to the relative disparity of these two measures.

Applying the algorithm in Table 1, column (i) to foreign-born adult (age 25 to 64) male workers in the 2000 US Census 1 percent PUMS data, the mean language proficiency score is 71.6 (standard deviation of 32.4). The mean for foreign-born adult male workers varies by occupation, from around 30.2 to 100.0 when the focus is restricted to occupations with 10 or more workers in the sample, and from 13.0 to 100.0 when occupations with smaller representation are also considered.[8] The mean English proficiency score using the values in Table 1, column (ii) is 57.6 (*e.g.*, taxi drivers and chauffeurs) and using the values of Table 1, column (iii) is 68.3 (*e.g.*, waiters and waitresses).

IV. EMPIRICAL ANALYSES

Given the information on the "required" level of English in each worker's occupation outlined in Section II, and that on the worker's actual English proficiency,

[8] Examples of occupations with these scores are: 13.0 - animal breeders; 30.2 - other extraction workers; 71.6 – hairdressers; 100.0 - financial examiners.

discussed in Section III, each worker can be assigned to relative English language categories as follows:

- $ENG_o = ENG_a - ENG_r$, if $ENG_a > ENG_r$,
  $= 0$, otherwise;
- $ENG_u = ENG_r - ENG_a$, if $ENG_r > ENG_a$, (1)
  $= 0$, otherwise.
- $ENG_a = ENG_r + ENG_o - ENG_u$;

where subscripts *a*, *o*, *u*, *r* designate the workers' actual proficiency, extent of overqualified language skills, extent of underqualified language skills and the "required" level of English in the respondent's occupation. The scores for ENG_r are obtained from either the O*NET database (see Section II) or the Realized Matches procedure (see Section V). The data for the actual English proficiency are the scores formed from the self-reported Census English proficiency question (see Table 1). ENG_o and ENG_u are non-negative numbers, and for any individual at most only one can be positive.

The augmented earnings function incorporating these additional relative measures may be termed the Over-, Required- and Under- Language (ORU-L) specification. It can be expressed as:

$$\ln Y_i = \beta X_i + \gamma_o ENG_{oi} + \gamma_r ENG_{ri} + \gamma_u ENG_{ui} + \eta_i \qquad (2)$$

where $\ln Y$ denotes the natural logarithm of earnings, *X* contains a set of standard determinants of earnings (including educational attainment, potential labor market experience, marital status, years of residence in the US and location), and η is a random error term. In the above specification, γ_o, γ_r and γ_u are the parameters to be estimated for ENG_o, ENG_r and ENG_u respectively. It is hypothesized that $\gamma_r > 0$, $\gamma_o > 0$ and $\gamma_u < 0$ and that $\gamma_r > \gamma_o$ and $\gamma_r > |\gamma_u|$. All variables used in the analyses are defined in Appendix A, which includes the means and standard deviations of the variables (Table A-1).

Table 2 presents estimates of the Occupational English Requirements models for the foreign born. The regression coefficients and t-ratios for the statistical control variables are similar to those generally found in the literature and in the interest of conserving space are not discussed here. The English requirements variables in Table 2 have been formed using the O*NET database, and the over/under mismatch variables have been formed using this information in combination with data on the workers' actual English proficiency scored according to column (i) of Table 1.[9] Two types of models are considered; a conventional earnings equation with an English requirement variable (Table 2, columns (i)), and an extended earnings equation with mismatch English qualification variables (columns (ii)). While estimates are presented based on both the O*NET level and the O*NET importance of English variables (Table 2), given the similarity of the findings, the discussions of results will focus only on the model with level of proficiency required for the occupation.[10] Estimates for the native born with only the required level and importance of English variables in the respondent's occupation are also presented, for comparison purposes.

The level of English variable captures the effect on earnings of the English requirements of the individual's occupation. There is a strong, positive relationship between earnings and the occupational English requirements, controlling for other variables, including years of schooling. An increase of one point in the standardized score of the level of English is associated with 1.7 percent higher earnings. In other words, across the range of hypothetical values of this variable (0 to 100), earnings differ by 1.7 log points. Comparing workers in occupations with standardized levels of English of 40 (*e.g.*, tax preparers and postal service clerks) and 20 (*e.g.*, glaziers and crossing guards),

[9] Estimates obtained using the other two methods for scoring workers' actual English proficiency detailed in Table 1 are presented and discussed in Section V.

[10] In these data, for the level of proficiency, 20 percent are correctly matched linguistically, while 68 percent are overqualified and 12 percent underqualified for their occupation. If the self-reported English proficiency data are adjusted to the same mean as the O*NET data, 26 percent of the workers are correctly matched, 42 percent are overqualified and 32 percent are underqualified. See Section V for discussion of this adjustment.

the implied earnings difference (of 0.34 log points) for immigrants is the equivalent of about eight years of schooling. [11] Occupational language requirements are very important to immigrants' labor market success.

English language requirements also have a substantial impact on the earnings of the native born (column (v)). The partial effect of the level of English variable for the native born is 0.013, though this is about 25 percent less than the effect of this variable among the foreign born (0.017).

The results in Table 2, column (ii) include the two English mismatch variables. These have been computed using equation (1). The inclusion of the relative overqualification and relative underqualification English variables in Table 2, column (ii) has minor impacts on the coefficients of most explanatory variables. The negative effect for earnings of being a Black immigrant increases from 9 percent to about 14 percent lower earnings than otherwise comparable non-Black immigrants when these mismatch variables are included.[12] The largest relative change, however, is for the educational attainment variable, where the estimate falls for the level of English analysis from 4.2 to 3.5 percent, or by 0.7 of a percentage point, a 19 percent drop.

The reduction in the partial effects of duration in the US (years since migration) when the over/under qualified language variables are added to the analysis (Table 2, column ii, compared to column i) is presumably due to the better matching (*i.e*. overqualified) of language skills and occupation with a longer stay in the US.

[11] This is much greater than when the effect of the respondent's English language skills is analyzed using a proficient/not proficient dichotomous measure. The effect of being proficient is the equivalent of about three additional years of schooling.

[12] Note that among the native born, controlling for either the level or importance of English in their occupation, Blacks have about 14 percent lower earnings than those of other races.

Table 2

Estimates of Earnings Functions With and Without the English Language Overqualification and Underqualification Variables, Foreign-Born and Native-Born Males, Aged 25-64, 2000 US Census

Variable	Foreign Born				Native Born	
	Level		Importance		Level	Importance
	(i)	(ii)	(iii)	(iv)	(v)	(vi)
Constant	4.982	4.952	5.038	5.021	4.233	4.191
	(93.69)	(92.40)	(94.43)	(92.98)	(190.10)	(187.72)
Years of Education	0.042	0.035	0.047	0.039	0.071	0.079
	(54.89)	(43.08)	(62.70)	(49.52)	(115.62)	(131.65)
Experience	0.013	0.015	0.011	0.014	0.035	0.035
	(12.39)	(14.82)	(10.79)	(13.39)	(77.12)	(76.80)
Experience Squared/100	-0.022	-0.024	-0.018	-0.021	-0.062	-0.061
	(11.03)	(12.23)	(9.37)	(10.69)	(61.57)	(60.75)
Married	0.203	0.194	0.209	0.199	0.257	0.261
	(34.55)	(33.18)	(35.42)	(33.87)	(107.20)	(108.41)
South	-0.080	-0.077	-0.079	-0.075	-0.064	-0.062
	(13.35)	(12.97)	(13.02)	(12.59)	(28.04)	(27.23)
Metropolitan Area	0.114	0.122	0.110	0.119	0.192	0.193
	(4.25)	(4.57)	(4.09)	(4.45)	(32.54)	(32.82)
Veteran	-0.038	-0.054	-0.048	-0.063	-0.037	-0.042
	(2.97)	(4.18)	(3.68)	(4.85)	(13.67)	(15.61)
Blacks	-0.086	-0.137	-0.099	-0.151	-0.132	-0.140
	(8.52)	(13.31)	(9.72)	(14.57)	(37.41)	(39.56)
Log Weeks Worked	0.878	0.869	0.880	0.871	1.003	1.004
	(73.63)	(72.90)	(73.74)	(72.96)	(183.11)	(183.13)
Years Since Migration (YSM)	0.015	0.011	0.015	0.010	(a)	(a)
	(18.31)	(13.34)	(17.65)	(12.44)		
YSM Squared/100	-0.014	-0.011	-0.012	-0.009	(a)	(a)
	(7.04)	(5.43)	(6.32)	(4.59)		
Required English (Level)	**0.017**	**0.019**	**(a)**	**(a)**	**0.013**	**(a)**
	(62.13)	**(67.91)**			**(112.44)**	
Required English (Importance)	**(a)**	**(a)**	**0.012**	**0.014**	**(a)**	**0.009**
			(52.74)	**(59.42)**		**(101.25)**
Overqualified English	**(a)**	**0.003**	**(a)**	**0.003**	**(a)**	**(a)**
		(20.86)		**(17.37)**		
Underqualified English	**(a)**	**-0.004**	**(a)**	**-0.004**	**(a)**	**(a)**
		(12.53)		**(16.75)**		
Adjusted R^2	0.381	0.388	0.372	0.380	0.352	0.349
Sample Size	84,172	84,172	84,172	84,172	531,821	531,821

Note: Heteroskedastic-consistent 't' statistics in parentheses.
Source: 2000 US Census, One-percent Public Use Microdata Sample.

The increased negative effect on earnings of being a Black immigrant when the language mismatch variables are included in Table 2, column ii, implies that Black immigrants are less well matched linguistically than others. Alternatively, deleting the quality of the match reduces their earnings disadvantage. It should be noted that Black immigrants come disproportionately from origins where English is the primary or an important second language, the English speaking parts of the Caribbean and Africa.[13]

The variable for relative overqualified English records the difference between the worker's English score and the level of English required in his job. The estimated coefficient is positive, at 0.003. This gain in earnings for levels of English proficiency in excess of that required in the job is much less than the 0.019 increase in earnings associated with levels of English proficiency required in the job. Thus, relatively overqualified workers gain some extra earnings for their surplus proficiency, but not as much as they would if they were to move to an occupation where their English skills are at the level required for the occupation.

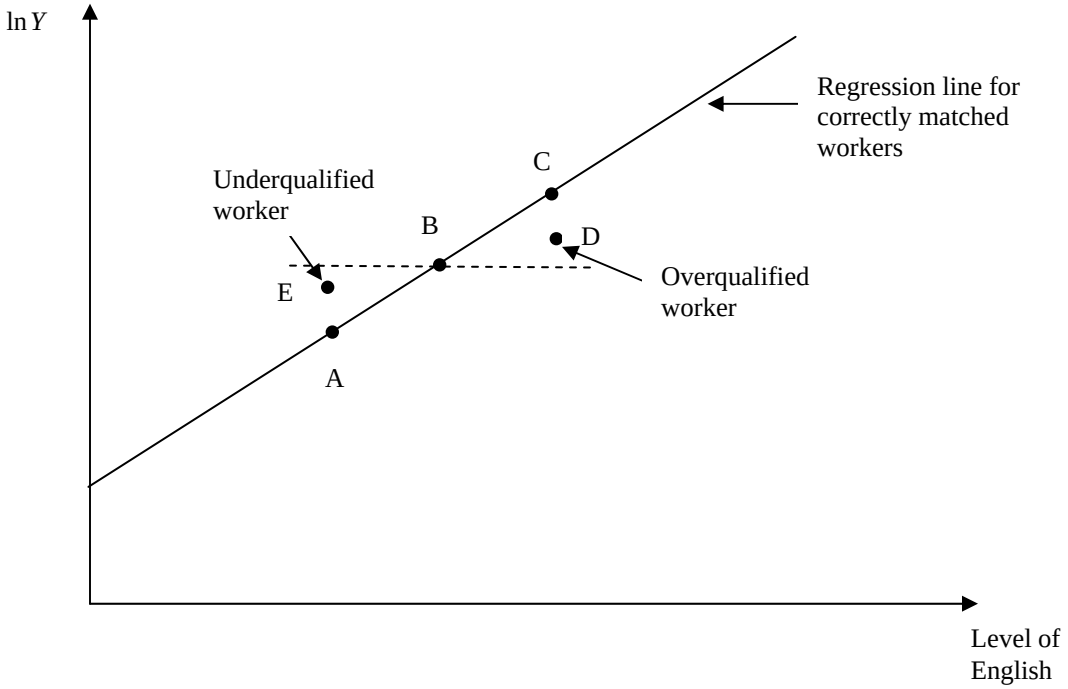
This pattern can be illustrated using hypothetical workers (Table 3 and Figure 3). Consider three workers, B, C and D (Table 3, Figure 3). Worker B has an English score of 50 and works in an occupation that requires exactly that level of English. That is, he is correctly matched in terms of his English proficiency. Worker C is also correctly matched in terms of his English skills, albeit at the higher score of 60 on the English scale. Worker D, however, is relatively overqualified in terms of his English skills, as he has an English proficiency score of 60 but is employed in an occupation that requires only the lower level of English of 50.

[13] In the 2000 Census, of the Black immigrants 31 percent were born in the English-speaking Caribbean Islands, 16 percent in the former British colonies of Africa, and 53 percent elsewhere.

Table 3
Earnings and Skill Classification of Five Hypothetical Workers

	Required Level of English	Actual Level of English	Skill Classification	ln*Y*
A	40	40	Correct Match	10.01
B	50	50	Correct Match	10.20
C	60	60	Correct Match	10.39
D	50	60	Overqualified	10.23
E	50	40	Underqualified	10.16

Figure 3
Earnings of Five Hypothetical Workers Described in Table 3



Comparing workers B and C, worker C will have 19 percent higher earnings than worker B (10 extra points on the occupational English requirements scale, valued at 1.9 percent higher earnings per point). Worker C gets this higher earnings for two reasons: first, he has a higher proficiency in English, and second, he moves to an occupation that is suited to his superior English skills. Thus, the 19 percent higher earnings can be

viewed as a payoff to the acquisition of a skill and inter-occupational mobility to where the skill can be used more effectively.

The increase in earnings can be decomposed into that due to higher earnings within an occupation and occupational mobility. Worker D has a higher English proficiency than worker B but is employed in the same occupation as worker B. That is, the inter-occupational mobility that characterizes worker C is missing in the case of worker D. Worker D earns 3 percent more than worker B, which is the increase in earnings associated with the higher proficiency in English. However, he earns 16 percent less than the correctly matched worker C, which is the earnings gain to workers from moving to an occupation suited to their superior English skills. Thus, of the 19 percent increase in earnings for worker C compared to worker B, 3 percentage points appear to be due to the acquisition of the greater English proficiency but remaining in the same occupation, and 16 percentage points appear to be due to worker C having moved to an occupation where the superior English skills can be used effectively.

The variable for a relatively underqualified English score has a negative coefficient, of –0.004 (Table 2, column ii). This indicates that the person will incur a negative return of 0.4 percent if his English score falls short of the job's required score by one point. The earnings outcome for underqualified workers is also best explained using an example (Table 3 and Figure 3). Consider worker D who has an actual English proficiency score of 40 who gains a position in an occupation that has an English requirement of 50 (Worker C). Compare this English underqualified worker to those in jobs with occupational English requirements of 40 and 50 who are correctly matched (Workers A and B). Given the specification of the earnings equation adopted here, compared to worker A, workers B and E receive the earnings increments associated with the higher English requirements of their occupation (*i.e.*, 10 extra points on the English requirements scale, valued at 0.019 per point, or 19 percent higher earnings). However, the fact that worker E is underqualified entails an earnings penalty compared to a correctly qualified worker in this occupation (worker B). This earnings penalty is -0.004 per point on the English scale, and so it is 4 percent for worker E. Thus, the

underqualified worker (worker E) in this comparison will earn 4 percent less than the correctly qualified worker B, who is employed in a job with the higher English requirement. Yet, worker E earns 15 percent more (*i.e.*, 19 percent for the higher English requirement less 4 percent for being underqualified) than the correctly qualified worker A who is employed in a job with the lower English requirement.

The estimates of the earnings effects associated with English language skills that are relatively overqualified, relatively underqualified, and for being correctly matched, mirror the findings from the undereducation/overeducation literature. This indicates that the central ideas of this earlier literature generalize to other forms of human capital. The findings presented above, however, may be sensitive to the way that the occupational English requirements and workers' English language proficiency are measured. These issues are examined in the following section.

V. ROBUSTNESS CHECKS

The sensitivity of the earnings effects associated with occupational English requirements and mismatched English skills to the measurement and specification of the English language variables is examined in this section. First, the sensitivity of the estimates to the way workers' English proficiency is scored (see Table 1) is examined. Then the sensitivity of the estimates to the measure of occupational English language requirements is investigated by replacing the O*NET measure (used to obtain the Table 2 estimates) with measures obtained using the observed Realized Matches methodology, based on the mean proficiency reported by the foreign born within occupations. Finally, the robustness of the findings to the use of a linear specification for the occupational English requirements is examined by generalizing the linear form to a quadratic.[14]

[14] A further robustness check of the sensitivity of the estimates to the way the measures of occupational English requirements and workers' English proficiency are combined is examined in Appendix C. This involves standardizing the distributions for these measures. This adjustment does not affect the general findings or material conclusions that might be drawn from the analysis.

A. Sensitivity to the Measure of Workers' English Proficiency

Table 4 lists estimates of the coefficients on the English language variables in the ORU specification for the three alternative ways of scoring a workers' English proficiency listed in Table 1. The general pattern of effects – in terms of sign, relative numerical magnitudes, and statistical significance – is not affected substantially by the choice of algorithm for scaling a worker's proficiency in English.

Table 4

Estimates for English Requirements and Mismatch Variables for Alternative Scores of Actual English Proficiency, Foreign-Born Males, Aged 25-64, 2000 US Census[(a)]

Variable	Level			Importance		
	(i)	(ii)	(iii)	(i)	(ii)	(iii)
Level/Importance of English	0.019 (67.91)	0.019 (68.68)	0.019 (67.65)	0.014 (59.42)	0.015 (60.35)	0.014 (59.19)
Overqualified English	0.003 (20.86)	0.003 (14.22)	0.003 (16.73)	0.003 (17.37)	0.002 (8.63)	0.003 (14.87)
Underqualified English	-0.004 (12.53)	-0.006 (21.17)	-0.005 (15.98)	-0.004 (16.75)	-0.006 (28.36)	-0.004 (18.84)
Adjusted R^2	0.388	0.389	0.389	0.380	0.382	0.380

Notes: (a) Each estimating equation contains the same set of standardizing variables as column (ii) of Table 2. Actual English scored according to Table 1, column (i) for column (i) here, column (ii) for column (ii) here and column (iii) for column (iii) here.

(b) Heteroskedasticity-consistent 't' statistics in parentheses.

Source: 2000 US Census, One percent Public Use Microdata Sample

B. Sensitivity to the Measure of Occupational English Requirements

The analyses reported above are based on occupational English requirements obtained from the O*NET database. As explained in Section II, this is a Worker Self-Assessment method for determining the level of English required in each occupation. An alternative approach is to use the Realized Matches method. This uses the self-reported English proficiency of workers in each occupation as the reference or usual level in the occupation. It is employed here to ascertain whether the findings in Table 2 are sensitive to the way the required level of English for each occupation is determined. This method actually represents the labor market outcome of the worker allocation process more so than the other measures.

In forming the Realized Matches measure, the Census categories for English proficiency (speaks only English at home; speaks a language other than English at home and speaks English "Very Well", "Well", "Not Well" or "Not at All") have been scored using the three algorithms for which values are listed in Table 1. Only the foreign born are used in the construction of these benchmarks.[15]

Comparisons between the estimations using the Realized Matches and O*NET measures can be enhanced by having distributions of workers across the overqualified, correctly qualified and underqualified categories which are similar for each of these measures. This can be achieved by subtracting a constant from the Realized Matches score for each occupation so that the mean of the values obtained under each of the three Realized Matches algorithms is the same as that for the O*NET data for the level of English required in the occupation (subject to the English requirement for each occupation being non-negative). This simple approach is taken to enable presentation of a set of analyses as comparable as possible to the results presented in Table 2.

Table 5 lists selected estimates from earnings equations based on the three different score-assignment frameworks discussed above.[16] The full set of results from this analysis based on the Realized Matches procedure is presented in Appendix B.

The results for these models are broadly similar, even though different algorithms for establishing the reference level of English are employed. Hence, the following discussion of results focuses only on the estimates labeled "Model I", which use the

[15] As the native born are almost all English-only speakers (about 95 percent) and most of the others report speaking English "very well", and they dominate the employment in the overwhelming majority of occupations, their inclusion in the sample used to form the required English levels would drive these benchmarks towards 100.

[16] Note that in each of these estimations, for internal consistency, the algorithm for computing the workers' proficiency in English score corresponds to that for compiling the reference level of English for the occupations.

scores for the English proficiency categories given in column (i) of Table 1 to establish the reference level of English for each occupation.

Table 5
Estimates for English Requirements and Variables Based on Realized Matches Procedure, Foreign-Born Males, Aged 25-64, 2000 US Census[(a)]

Variable	Model I	Model II	Model III
English Requirement, Realized Matches Procedure	0.018 (78.79)	0.021 (79.16)	0.018 (79.56)
Overqualified English	0.002 (11.88)	0.001 (8.22)	0.001 (9.07)
Underqualified English	-0.005 (14.06)	-0.006 (18.92)	-0.006 (17.06)
Adjusted R^2	0.397	0.398	0.398
Sample Size	84,172	84,172	84,172

Notes: Heteroskedastic-consistent 't' statistics in parentheses. Each estimating equation contains the same set of standardizing variable as column (ii) of Table.

(a) English requirements based on actual proficiency in English measures with scores from Table 1 column (i) used in Model I, column (ii) used in Model II, and column (iii) in Model III.

The full regressions are reported in Appendix B.

Source: 2000 US Census, , One percent Public Use Microdata Sample.

The English requirement variable in the first of these alternative analyses, which gives the return to worker's correctly-matched English skills, has a positive coefficient of 0.018, and a 't' value of 80. This compares with the coefficient of 0.019 and 't' value of 68 reported for the O*NET level of English variable in Table 2.

The English mismatch variables in Table 5 are associated with conventional results. The overqualification variable implies that positive earnings of 0.2 percent (t = 11.9) are associated with each point score in excess of that required in the job. In the analyses reported in Table 2, based on the O*NET database, the earnings increment associated with overqualified English skills was 0.3 percent (t = 20.9). The underqualification variable, which captures the penalty for each score that falls short of that required by the job, has a negative coefficient of -0.005 (t = 14.1) in Table 5. This is

similar in value as that (-0.004, t = -12.5) estimated using the O*NET Level of English information.

Hence, the point estimates for the earnings effects differ only slightly, and the general pattern of effects for the required and mismatch English variables is the same across the various measures (Worker Self-Assessment from O*NET or Realized Matches from the Census) of occupational English requirements considered.[17]

C. Sensitivity to the Linear Specification of the English Language Requirements Variable

Table 6 contains results from the ORU-Language (ORU-L) earnings equation with the O*NET occupational English language requirements entered as a quadratic.[18] For comparative purposes, the results from column (ii) of Table 2 are also presented.

These results show that the earnings returns associated with occupational English language requirements increase at an increasing rate. For the level of English, the partial effect changes from around 1.5 percent at a level of English of 20, to around 2.3 percent at a level of English of 80 (compared to the uniform partial effect of 1.9 percent in the linear specification). Similarly, for the importance of English, the partial effect changes from 1.00 percent at an importance score of 20, to 1.82 percent at an importance score of 80 (compared to the uniform effect of 1.4 percent in the linear specification). The coefficients on the overqualification in English and underqualification in English variables, however, are the same as reported previously on the basis of the linear specification. Moreover, the statistical significance of the mismatch variables is unaltered by making the English language requirements variable quadratic.

[17] Hartog (2000) comes to similar conclusions regarding the robustness of the results when the technique is applied to education.

[18] Hartog (2000, p.135) performs a similar test in his analysis for education. Because of the smaller variation in the measures of overqualified and underqualified English language skills, quadratic terms cannot be entered for these variables.

Table 6
Partial Effects of Language Variables with Quadratic Level/Importance Variables, Foreign-Born Males, Aged 25-64, 2000 US Census

Variable	Level		Importance	
	(i)	(ii)	(i)	(ii)
Required English (Level)	0.019 (67.91)	0.012 (9.05)	(a)	(a)
Level of English Squared/100	(a)	0.007 (4.74)	(a)	(a)
Required English (Importance)	(a)	(a)	0.014 (59.42)	0.007 (5.57)
Importance of English Squared/100	(a)	(a)	(a)	0.007 (6.22)
Overqualified English	0.003 (20.86)	0.003 (21.24)	0.003 (17.37)	0.003 (17.77)
Underqualified English	-0.004 (12.53)	-0.004 (12.10)	-0.004 (16.75)	-0.004 (16.26)
Adjusted R^2	0.388	0.388	0.380	0.380
Sample Size	84,172	84,172	84,172	84,172

Note: Heteroskedastic-consistent 't' statistics in parentheses. Each equation includes the same set of standardizing variables as in column (ii) of Table 2.
Source: 2000 US Census, One percent Public Use Microdata Sample.

VI. SUMMARY AND CONCLUSION

This paper examines the effect on earnings of the mismatch of a male immigrant worker's English language proficiency and the proficiency required in his occupation. The empirical analyses in this paper use data on foreign-born and native-born adult (age 25 to 64) men from the 2000 US Census (one percent public use microdata sample) and information on occupational English language requirements obtained from the O*NET database (version 12.0).

Workers are identified whose English language skills were correctly matched to their job requirements, those who were overqualified, and those who were underqualified in terms of their English skills. The earnings consequences are computed for the correct matching and the mismatch of a worker's English language proficiency to the requirements (level and importance) of the occupation in which he works.

The results obtained using the occupational English requirements specified in the O*NET database suggest that earnings are positively and significantly associated with occupational English language requirements. These earnings gains are shown to arise for two reasons. First, there is a positive, although relatively minor, payoff to the acquisition of English language skills while remaining within one's occupation. Second, there is a more important payoff to mobility to an occupation better suited to the higher English skills. Thus, overqualified workers earn modest rewards for their excess endowment of English skills, but these rewards are far less than what they would receive if they moved to jobs in which they were linguistically correctly matched. Underqualified workers incur earnings penalties for their skill inadequacies, compared to workers in their occupation who have the right level of English skills for their job.

Several tests were conducted of the sensitivity of the effects on earnings to the measure of workers' English proficiency, the measure of occupational English requirements, the distribution of the variables, and to the functional form (linear or quadratic) used for the occupational English language requirements variable. The tests confirmed the robustness of the estimates.

The results from this paper have demonstrated the fruitfulness of the ORU (Over/Required/Under qualified) approach to the study of language proficiency. Future research could seek to establish whether these empirical regularities regarding language skills, occupation and earnings based on the 2000 Census carry across to analyses of other data sets for the US and to the study of the labor markets of other countries. Given the strength of the findings, consideration might also be given to examining whether the earnings consequences of other forms of human capital (*e.g.*, health capital, computer skills) can be quantified using the under/over skill matching framework. As skill mismatches suggest sub-optimal and inefficient resource allocations, policy actions might encourage better job matching in the labor market for education and language skills. This could be an important role for immigrant settlement or assimilation policies.

RERERENCES

Bleakley, Hoyt and Chin, Aimee, (2004). "Language Skills and Earnings: Evidence from Childhood Immigrants", *Review of Economics and Statistics*, 86(2), pp.481-496.

Chiswick, Barry R. and Miller, Paul W., (1995). "The Endogeneity Between Language and Earnings: International Analyses", *Journal of Labor Economics*, 13(2), pp. 246-288.

Chiswick, Barry R. and Miller, Paul W., (1998a). *The Economic Cost to Native-Born Americans of Limited English Language Proficiency*, Report prepared for the Center for Equal Opportunity, August 1998.

Chiswick, Barry R. and Miller, Paul W., (1998b). "Census Language Questions in North America", *Journal of Economic and Social Measurement*, 25(2), pp.73-95.

Chiswick, Barry R. and Miller, Paul W., (2007) *The Economics of Language: International Analyses* Routledge.

Chiswick, Barry R. and Miller, Paul W., (2008a). "Modeling Immigrants' Language Skills", *Research in Labor Economics*, Vol. 27, pp. 75-128.

Chiswick, Barry R. and Miller, Paul W., (2008b). "Why is the Payoff to Schooling Smaller for Immigrants?", *Labour Economics*, Vol. 15, No. 6, December, pp. 1317–1340.

Chiswick, Barry R. and Miller, Paul W., (2010). "Occupational Language Requirements and the Value of English in the United States Labor Market" *Journal of Population Economics*, Vol. 23, No. 1, pp. 353-372.

Duncan, Greg and Hoffman, Saul D., (1981). "The Incidence and Wage Effects of Overeducation", *Economics of Education Review*, 1(1), pp.75-86.

Espenshade, Thomas J. and Fu, Haishan, (1997). "An Analysis of English-Language Proficiency Among U.S. Immigrants", *American Sociological Review*, 62(2), pp.288-305.

Evans, Mariah D.R., (1986). "Sources of Immigrants' Language Proficiency: Australian Results with Comparisons to the Federal Republic of Germany and the United States of America", *European Sociological Review*, 2(3), pp.226-236.

Evans, Mariah D.R., (1987). "Language Skill, Language Usage and Opportunity: Immigrants in the Australian Labour Market", *Sociology*, 21(2), pp.253-274.

Hartog, Joop, (2000). "Over-Education and Earnings: Where Are We, Where Should We Go?", *Economics of Education Review*, 19(2), pp.131-147.

Kominski, Robert, (1989). “How Good is “How Well”? An Examination of the Census English-Speaking Ability Question”, *American Statistical Association 1989 Proceedings of the Social Statistics Section*, pp.333-338.

Rumberger, Russell W., (1987). “The Impact of Surplus Schooling on Productivity and Earnings”, *Journal of Human Resources*, 22(1), pp.24-50.

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 United States Census of Population, 1 percent Public Use Microdata Sample; O*NET Occupational data (version 12.0) from the O*NET Consortium described in Section II.

Definition of Population: Foreign-born (and native-born) men aged twenty-five to sixty-four who worked at least one week in 1999. Only residents of the 50 States and the District of Columbia are considered.

Dependent Variable:

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

Explanatory Variables:

Educational Attainment (EDUC): This variable records the total years of full-time equivalent education. It has been constructed from the Census data on educational attainment by assigning the following values 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 one year (12.5); attended college for more than one year or completed college (14); Bachelor's degree (16); Master's degree (17.5); Professional degree (18.5); Doctorate (20).

Labor Market Experience (EXP): This is a measure of potential labor market experience, computed as AGE – Years of Education – 6.

Weeks Worked (WEEKS): The number of weeks the individual worked in 1999 is entered into the specification in natural logarithmic form.

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): This is 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 of a non-metropolitan area (NON-MET) or of the Southern States (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 a veteran of the US Armed Forces, and set to zero otherwise.

Years Since Migration (YSM). This is computed from the year the foreign-born person came to the United States to stay.

Actual English Language Proficiency: This is a continuous variable formed from the Census information on whether the individuals (i) speak only English at home, or speak a language other than English in the home and speak English either: (ii) “Very Well”; (iii) “Well”; (iv) “Not Well”’ (v) “Not at All”. The values assigned to these categories are listed in Table 1 and the specific variable used in each estimation is noted in the text.

English Requirements (LEVEL and IMPORTANCE): These variables record the scores for the level and importance of English requirements for each occupation code obtained from the O*NET database (http://online.onetcenter.org). Tests of robustness are

conducted using a Realized Matches procedure where the reference level of English for each occupation is given by the mean actual English proficiency of foreign born workers in each occupation.

English Overqualification (OVERQUALIFIED): The overqualification variable equals the difference between the worker's actual score for English proficiency and the English score required by the job, where this computation is positive. Otherwise, it is set equal to zero.

English Underqualification (UNDERQUALIFIED): The underqualification variable equals the difference between the English score required by the worker's job and the worker's actual score for English proficiency, where this computation is positive. Otherwise, it is set equal to zero.

Table A-1
Measured Standard Deviations for Variables used in Estimating Equations
Foreign and Native Born Males, Age 25-64, 2000 Census

Variable	Foreign Born	Native Born
Log Income	10.122 (0.99)	10.379 (0.98)
Years of Education	11.871 (4.78)	13.663 (2.51)
Experience	22.224 (10.91)	22.423 (10.44)
Experience Squared/100	612.808 (558.98)	611.737 (509.02)
Married	0.645 (0.48)	0.654 (0.48)
South	0.281 (0.45)	0.361 (0.48)
Metropolitan Area	0.990 (0.10)	0.960 (0.20)
Veteran	0.048 (0.21)	0.237 (0.43)
Blacks	0.075 (0.26)	0.102 (0.30)
Log Weeks Worked	3.766 (0.47)	3.820 (0.41)
Years Since Migration (YSM)	16.620 (10.95)	-
YSM Squared/100	396.122 (478.20)	-
English Requirement, Level of English	45.691 (12.28)	48.259 (12.21)
English Requirement, Importance of English	55.647 (14.28)	58.767 (14.56)
Overqualified English (Level)	30.194 (23.17)	-
Underqualified English (Level)	4.254 (10.68)	-
Overqualified English (Importance)	22.791 (20.45)	-
Underqualified English (Importance)	6.808 (14.02)	-
Sample Size	84,172	531,821

Note: Standard deviations in parentheses.
Source: 2000 US Census, One percent Public Use Microdata Sample

APPENDIX B
Estimates of Earnings Function With Overqualification and Underqualification Variables Formed Using Realized Matches Procedure, Foreign-Born Males, Aged 25-64, 2000 US Census[(a)]

Variable	Model I	Model II	Model III
Constant	5.403	5.315	5.443
	(103.47)	(102.01)	(104.42)
Years of Education	0.027	0.026	0.026
	(31.56)	(31.59)	(30.72)
Experience	0.016	0.016	0.016
	(15.78)	(15.54)	(16.03)
Experience Squared/100	-0.027	-0.026	-0.027
	(13.85)	(13.57)	(14.05)
Married	0.183	0.184	0.182
	(31.53)	(31.83)	(31.44)
South	-0.074	-0.074	-0.073
	(12.47)	(12.49)	(12.44)
Metropolitan Area	0.057	0.064	0.060
	(2.15)	(2.42)	(2.26)
Log Weeks Worked	0.855	0.855	0.855
	(72.18)	(72.20)	(72.19)
Veteran	-0.083	-0.087	-0.083
	(6.46)	(6.76)	(6.46)
Blacks	-0.149	-0.167	-0.151
	(14.65)	(16.06)	(14.76)
Years Since Migration (YSM)	0.009	0.010	0.009
	(11.18)	(11.71)	(11.04)
YSM Squared/100	-0.007	-0.009	-0.007
	(3.46)	(4.44)	(3.48)
English Requirement, Realized Matches Procedure	**0.018**	**0.021**	**0.018**
	(79.18)	**(79.56)**	**(80.08)**
Overqualified English	**0.002**	**0.002**	**0.002**
	(14.68)	**(13.10)**	**(12.25)**
Underqualified English	**-0.005**	**-0.006**	**-0.006**
	(10.73)	**(15.54)**	**(15.27)**
Adjusted R^2	0.397	0.398	0.398
Sample Size	84,172	84,172	84,172

Notes: Heteroskedastic-consistent 't' statistics in parentheses.

(a) When forming the occupational English requirements and workers' actual proficiency in English measures, the scores from Table 1 columns i, ii, and iii were used here for columns i, ii, and iii respectively.

Source: 2000 US Census, One percent Public Use Microdata Sample.

APPENDIX C

SENSITIVITY TO THE DISTRIBUTIONS OF THE VARIABLES

Each of the benchmark English requirements variables and the scores of workers' actual English proficiency is recorded using different scales of measurement. This gives rise to different means and standard deviations. To enhance the comparability of the findings of the analysis based on the O*NET information on level of English and the information compiled using the Realized Matches procedure, the Realized Matches variables were scaled in Section V so that they had the same mean as the O*NET level of English variable.

In the analyses below this theme is developed by converting all English measures, both of the occupational English requirements and of the workers' English proficiency, to standardized scores. These standardized scores have a zero mean and a unit variance. Selected results are reported in Table C.1.

Table C.1
Estimates for English Requirements and Mismatch Variables Using Standardized Measures of English Requirements and Proficiency, Foreign-Born Males, Aged 25-64, 2000 US Census

Variable	Occupational English Requirements			
	O*NET Level of English[(a)]	Realized Matches[(b)]		
		I	II	III
Required Level of English	0.291 (66.40)	0.317 (73.58)	0.328 (75.61)	0.321 (74.72)
Overqualified English	0.043 (8.13)	0.040 (7.83)	0.041 (8.29)	0.036 (7.06)
Underqualified English	-0.156 (29.07	-0.109 (18.65)	-0.130 (20.95)	-0.122 (20.30)
$\overline{R}^2$	0.390	0.397	0.398	0.398
Sample Size	84,172	84,172	84,172	84,172

Notes: For notes on the Realized Matches variables, see Table 5.
(a) The workers' actual proficiency in English is scored using the algorithm of column (i) in Table 1.
(b) The algorithm for computing the workers' proficiency in English score (see Table 1) corresponds to that for compiling the reference level of English in these estimations.
Source: 2000 US Census, One percent Public Use Microdata Sample.

The estimates from these models are reassuring. The pattern of effects is the same for each specification of the English variables. Moreover, there is reasonably limited variation in the estimated coefficients for the English requirements and overqualified English variables. The estimated impact for the required level of English variable varies from 0.291 to 0.328, while the effect of the overqualified English variable varies from 0.036 to 0.043. The algorithms that results in relatively high estimated impacts for the required level of English are also associated with relatively high estimated impacts for overqualified English, although the relationship is not exact.

There is more variation across the alternative standardized measures in the estimated effects of underqualified English, with the point estimates ranging from -0.109

to -0.156. The unstandardized data analyzed previously were also characterized by greater variability in the earnings penalty associated with underqualified English than with the other two English variables in the ORU specification of the earnings equation. Moreover, the relative magnitudes of the differences in estimated effects across the algorithms used for determining the occupational English language requirements is affected by the use of standardized data. Thus, while the reference level obtained straight from the O*NET database was associated with relatively small (in absolute value) estimated effects of underqualified English (compare column (ii) of Table 2 with Table 5), the estimated effects of underqualified English in Table 6 for the standardized O*NET data are relatively large (again in absolute value). This suggests that the scale of measurement (the effects of which are neutralized in Table C.1) matters, however, only in terms of the point estimates, and not in terms of the general findings or material conclusions that might be drawn from the analysis. This issue does not appear to have been addressed in the overeducation/undereducation literature.