ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Chiswick, Barry R.; Miller, Paul W.

Working Paper

Matching language proficiency to occupation: the effect on immigrants' earnings

IZA Discussion Papers, No. 2587

Provided in Cooperation with:
IZA – Institute of Labor Economics

Suggested Citation: Chiswick, Barry R.; Miller, Paul W. (2007) : Matching language proficiency to occupation: the effect on immigrants' earnings, IZA Discussion Papers, No. 2587, Institute for the Study of Labor (IZA), Bonn, https://nbn-resolving.de/urn:nbn:de:101:1-2008042579

This Version is available at:
https://hdl.handle.net/10419/33960

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

DISCUSSION PAPER SERIES

IZA DP No. 2587

Matching Language Proficiency to Occupation: The Effect on Immigrants' Earnings

Barry R. Chiswick
Paul W. Miller

January 2007

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Matching Language Proficiency to Occupation: The Effect on Immigrants' Earnings

Barry R. Chiswick
University of Illinois at Chicago
and IZA

Paul W. Miller
University of Western Australia
and IZA

Discussion Paper No. 2587
January 2007

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of the institute. Research disseminated by IZA may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit company supported by Deutsche Post World Net. The center is associated with the University of Bonn and offers a stimulating research environment through its research networks, research support, and visitors and doctoral programs. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

IZA Discussion Paper No. 2587
January 2007

ABSTRACT

Matching Language Proficiency to Occupation: The Effect on Immigrants' Earnings*

This paper analyzes the effect on earnings of the matching of English language skills to occupational requirements. 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. The analyses show that earnings are related to correct matching of an individual's language skills and that of his occupation. Moreover, the findings are robust with respect to a range of measurement and specification issues. Immigrant settlement policy may have a role to play in matching immigrants to jobs that use their language skills most effectively.

JEL Classification: J24, J31, F22

Keywords: English language, earnings, immigrants, schooling

Corresponding author:

Barry R. Chiswick
Department of Economics
University of Illinois at Chicago
601 South Morgan Street
Chicago, IL 60607-7121
USA
E-mail: brchis@uic.edu

* We thank Derby Voon for research assistance. Chiswick acknowledges research support from the Institute of Government and Public Affairs, University of Illinois. Miller acknowledges financial assistance from the Australian Research Council.

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

I. INTRODUCTION

The undereducation/overeducation literature focuses on how a worker's level of education matches the level required in his or her job (see Hartog, 2000). This literature distinguishes three categories of workers. These are workers whose actual level of education matches the reference or required level for their job (termed "correctly matched"), those whose actual level of education is less than the reference level for their occupation (termed "undereducated"), and workers whose actual level of education is greater than the reference level (termed "overeducated"). Wages have been shown in a large number of countries, including the US, to vary according to which of these categories the worker is in (see Duncan and Hoffman, 1981; Rumberger, 1987).

There are other human capital skills to which the methodology of the undereducation/overeducation literature could be applied. One such 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. To date, however, the research in this area has not examined whether the earnings premium to language proficiency varies according to the skill requirements of the jobs in which immigrants work.[1]

This paper examines whether the central premise of the undereducation/overeducation literature carries across to the study of language skills. The analysis commences in Section II with a discussion of the language requirements of occupations. Ways of relating information on an individual's self-reported language proficiency to the language requirement of their occupation are assessed in Section III, with the main issues being

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

illustrated using data from the 2000 US Census, one percent microdata file. Section IV presents an analysis of earnings, with a focus on the matching of language skills to occupation. As English proficiency is nearly universal among the native born, these analyses are limited to foreign-born males. A number of robustness checks are considered in Section V. Section VI offers a conclusion.

II. OCCUPATIONAL ENGLISH LANGUAGE REQUIREMENTS[2]

The O*NET Database

Official data sources, such as the Occupational Information Network, or O*NET, database, record the level of English required in a range of occupations.[3] 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 in 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

[2] Parts of this presentation are from Chiswick and Miller (2006a), 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 undereducation/overeducation perspective.

[3] The National O*NET Consortium was organized to accomplish the development of O*NET and its related products for the US Department of Labor, Employment and Training Administration.

who did not feel that English was important to the performance of their current job were coded as zero. 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. Hence, 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.

When the O*NET database was first set up, 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 businesses). These survey data have been progressively integrated into the initial O*NET database. About one-half of occupations have new survey information in version 8 of the database used in this study. Hence, the job requirements obtained from the O*NET database should be viewed as having been compiled using a hybrid of the Job Analyst and Worker Self-Assessment approaches identified in the undereducation/overeducation literature.

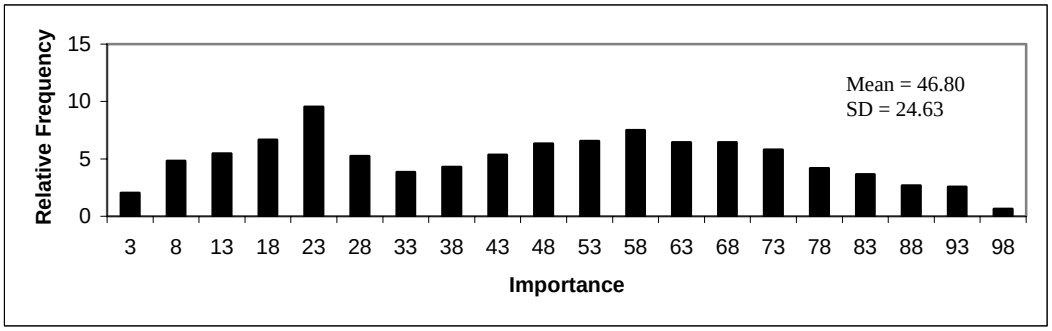
Importance of English

There is considerable variation in the importance of knowledge of the English language to job performance. The mean standardized score is 46.8, which is close to the mid-point ("Important") of the scale used in data collection. The standard deviation is 24.6, which is the equivalent of a change of one category on the underlying five-point scale. Moreover, the importance varies from minimal amounts (standardized scores around zero) in some occupations, to occupations where knowledge of English is very important.

Occupations where English is not important include “Maids and Housekeeping Cleaners” (score of 0), “Bakers, Bread and Pastry” (4) and “Slaughterers and Meat Packers” (8). Examples of occupations where English is very important are “Public Relation Specialists” (90), “Government Service Workers” (92) and “Judges, Magistrate Judges and Magistrates” (95).

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.[4, 5]

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



Level of English

The data on the level of English for all occupations are illustrated in Figure 2. The mean for all occupations is 39.2, which is midway 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 19. The occupations in the US labor market cover a wide range of the standardized scores, although there is limited representation above scores of 85. Compared to the

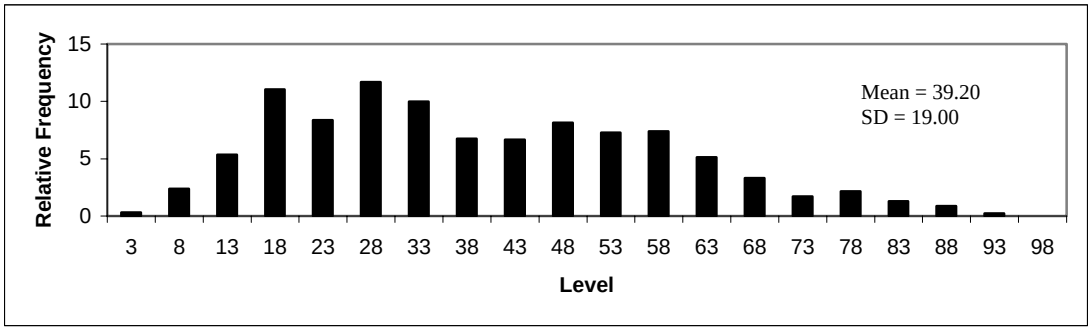
[4] The 933 occupations for which details are available in the O*NET database are used in the compilation of the data for this figure. They were not weighted by the number of workers in the occupation, and hence the distribution is sensitive to how jobs are grouped into occupations.

[5] For Figure 1 and related relative frequency distributions, 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.

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 one-third of the standardized scores. Nevertheless, there is a very high correlation across occupations between the scores for the importance of English and the level of English: it is 0.93 (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



Robustness Check

The empirical work 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 score for the English language requirements of each occupation given by the mean English proficiency of incumbents. 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. They provide a reference point for assessing whether workers have "too much" English for the job, "too little" English, or the right amount of English.

Information on the English language skills of workers is collected in many surveys, and these surveys 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 English proficiency to be categorized using a five-interval scale. The first 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. There has been debate 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 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. The first of these follows Evans (1986)(1987), and assigns the values 0, 33, 67 and 100 to the four English proficiency categories that she used. Evans (1986)(1987) 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. The basis for this particular scoring was that analyses of occupational status, the dependent variable in Evans' studies, that used categorical English proficiency variables showed that the effects of English proficiency on occupational attainment were approximately linear.

The second separates the English "Very Well" and "English only group" and treats the English only group as having superior skills to those who speak a language other than English at home and speak English very well. 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".[6] Within this five-category setting, values of 0, 25, 50, 75 and 100 are assigned to the English proficiency categories, 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 five English proficiency categories were examined, and these differences were used to establish an alternative ranking among the five English proficiency categories. This ranking 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 proficiency in English information.

Table 1
Values Assigned to Census Proficiency in English Information

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

Applying the algorithm in Table 1, column (i) to adult foreign-born adult male workers in the 2000 US Census 1 percent PUMS data, the mean language proficiency score is 71.6

[6] Chiswick and Miller (2007), however, 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.

(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. The mean English proficiency score using the algorithm in Table 1, column (ii) is 57.6 and that using the method of Table 1, column (iii) is 68.3.

IV. EMPIRICAL ANALYSES

Given the information on the required level of English in each worker's job outlined in Section II, and that on the worker's actual English proficiency, discussed in Section III, each worker can be assigned to categories akin to those used in the overeducation/undereducation literature 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 ENG_o denotes the extent of overqualification in terms of English skills, ENG_u denotes the extent of underqualification in terms of English skills, ENG_r is the required level of English in the occupation, and ENG_a is the worker's score on the actual English proficiency scale. 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 continuous score formed from the self-reported Census English proficiency data (see Table 1). ENG_o and ENG_u are non-negative numbers, and at most only one can be positive.

The augmented earnings function incorporating these additional measures may be termed the Over-, Required and Under- English (ORU) specification.[7] 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, controlling for the level or importance of English (both of which are expected to have a positive effect on earnings), $\gamma_o > 0$ and $\gamma_u < 0$. All variables used in the analyses are defined in Appendix A.

Table 2 presents estimates of the Occupational English Requirements models for the foreign born. Two types of models are considered; a conventional earnings equation with an English requirement variable (columns (i)), and an extended earnings equation with mismatch English qualification variables (columns (ii)). The English requirements variables have been formed using the O*NET database, and the mismatch variables have been formed using this information in combination with data on workers' actual English proficiency scored according to column (i) of Table 1.[8] While estimates are presented based on both the O*NET level and the importance of English variables (Table 2, columns (iii) and (iv)), given the similarity of the findings, the discussions of results will focus only on the model with level of English.

[7] This study follows Hartog (2000) in naming this earnings function as the ORU specification.

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

Table 2
Estimates of Earnings Function With and Without the English Language Overqualification and Underqualification Variables, Foreign-Born Males Aged 25-64, 2000 US Census

Variable	Level		Importance	
	(i)	(ii)	(i)	(ii)
Constant	5.366	5.327	5.404	5.394
	(101.89)	(100.83)	(102.72)	(102.10)
Log Weeks Worked	0.876	0.868	0.876	0.867
	(73.36)	(72.68)	(73.36)	(72.66)
Years of Education	0.044	0.037	0.046	0.038
	(54.17)	(43.93)	(57.97)	(46.86)
Experience	0.012	0.015	0.011	0.013
	(12.00)	(14.31)	(10.64)	(13.09)
Experience Squared/100	-0.022	-0.024	-0.019	-0.021
	(10.98)	(12.01)	(9.62)	(10.76)
Married	0.203	0.194	0.205	0.195
	(34.37)	(32.97)	(34.60)	(33.03)
South	-0.072	-0.070	-0.072	-0.069
	(12.03)	(11.68)	(11.90)	(11.55)
Metropolitan Area	0.099	0.108	0.096	0.106
	(3.68)	(4.04)	(3.57)	(3.96)
Veteran	-0.047	-0.061	-0.052	-0.065
	(3.64)	(4.72)	(4.04)	(4.97)
Blacks	-0.102	-0.150	-0.106	-0.153
	(10.02)	(14.55)	(10.43)	(14.78)
Years Since Migration (YSM)	0.014	0.010	0.014	0.010
	(17.17)	(12.25)	(16.91)	(11.84)
YSM Squared/100	-0.011	-0.008	-0.011	-0.007
	(5.72)	(4.10)	(5.56)	(3.80)
Level of English	0.011	0.013	(a)	(a)
	(53.52)	(60.86)		
Importance of English	(a)	(a)	0.008	0.010
			(51.18)	(58.72)
Overqualified English	(a)	0.003	(a)	0.002
		(21.86)		(17.08)
Underqualified English	(a)	-0.005	(a)	-0.006
		(12.18)		(18.45)
Adjusted R^2	0.373	0.380	0.371	0.378
Sample Size	84,172	84,172	84,172	84,172

Note: Heteroskedastic-consistent 't' statistics in parentheses.
Source: 2000 US Census, 1% PUMS.

According to the results in Table 2, column (i), the payoff to an additional year of education for foreign-born males in the US labor market is around 4.5 percent. This payoff is quite modest compared to the payoff of approximately 10 percent usually reported for the native born, but this is a typical pattern. The foreign born also have comparatively low payoffs to pre-immigration labor market experience. Thus, the payoff to a year of labor market experience accumulated in the country of origin is only 0.8 percent (evaluated at 10 years of experience). The separate effect of an extra year's duration of residence is around 1.2 percent, when evaluated at 10 years of residence. The addition of the pre-immigration labor market experience and duration of residence coefficients gives the impact on earnings of post-immigration labor market experience. This is around 2 percent at 10 years of experience, a payoff only slightly less than the 2.3 percent for the native born reported in Chiswick and Miller (2006b).

The natural logarithm of weeks worked in 1999 is a continuous variable that indicates an elasticity of earnings with respect to weeks worked of around 0.88. Other individual characteristic variables that have significant effects on the earnings of foreign-born males are marital status, South, metropolitan area, veteran of the US Armed Forces, and race. According to these estimates, married men earn 20 percent more than their unmarried counterparts, living in a metropolitan area is associated with around 10 percent higher earnings, while living in the southern states, having been a veteran of the US Armed Forces and having a Black racial origin have negative impacts of around 7 percent, 4 percent and 10 percent, respectively. These estimated impacts are comparable to those reported in the recent literature, even though the specification of the earnings equation used here includes a variable for the occupational English language requirements.

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. An increase of one point in the standardized score of the level of English is associated with 1.1 percent higher earnings. In other words, across the range of hypothetical values of this variable (0 to 100), earnings differ by 1.1 log points. This is the equivalent of a change in educational attainment of 24

years.[9] Comparing workers in occupations with standardized levels of English of 40 (*e.g.*, **Derby..we need an example**) and 20 (*e.g.*, earth drillers and tool and die makers), the implied earnings difference (of 0.22 log points) is the equivalent of about five years of schooling. Occupational language requirements are very important to immigrants' labor market success.

The results in Table 2, column (ii) include the two English mismatch variables. These have been computed using equation (1). The inclusion of the overqualification and underqualification English variables in Table 2, column (ii) has minor impacts on other coefficients, with the largest relative change being for the educational attainment variable, where the estimate falls by 0.7 of a percentage point, a 16 percent drop. An implication of this change is that additional years of schooling are associated with improved matching of worker's language skills to occupational requirements.

The variable for overqualified English records the difference between the individual'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 0.013 increase in earnings associated with levels of English proficiency required in the job. Thus, overqualified workers gain some extra earnings for their surplus skills, but not as much as they would if they were able to move to an occupation where their English skills were fully utilized.

This pattern can be illustrated using hypothetical workers. Consider three workers, B, C and D. Worker B has an English score of 40 and works in an occupation that requires exactly that level of English. That is, he is correctly matched in terms of his English skills. Worker C is also correctly matched in terms of his English skills, albeit at the higher score of 50 on the English scale. Worker D, however, is overqualified in terms of his English skills, as he has an English proficiency score of 50 but is employed in an

[9] This is much greater than when the effect of 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.

occupation that requires only the lower level of English of 40. The earnings and skill classification of these workers, together with several others to be discussed later, are summarized in Table 3, and their relative earnings positions displayed in Figure 3.

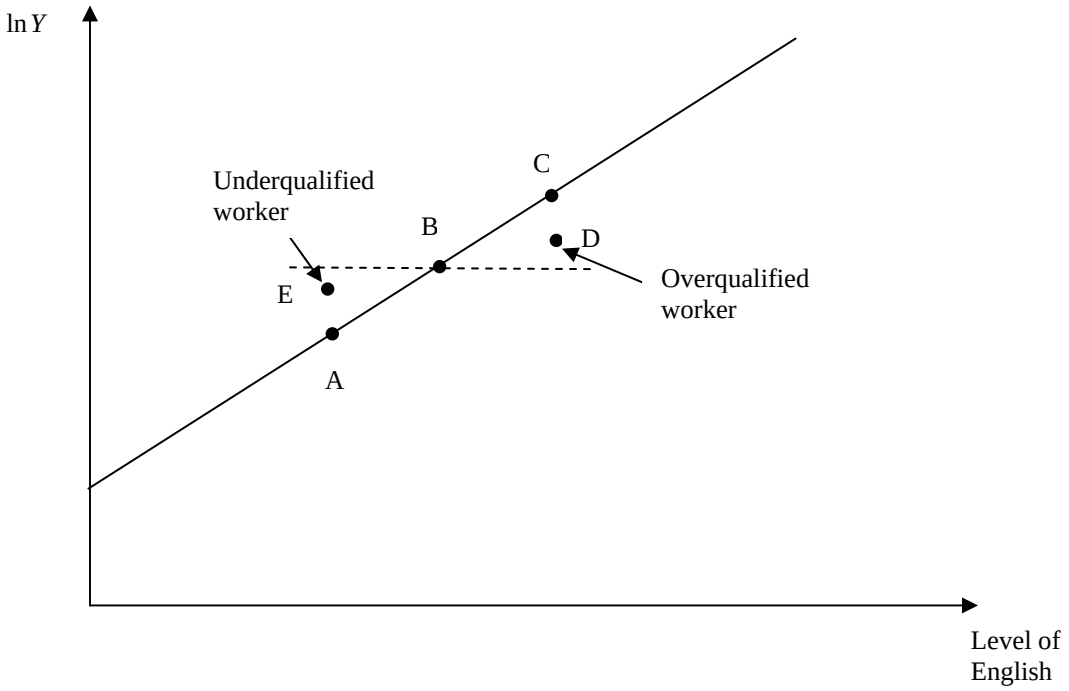
Comparing workers B and C, worker C will have 13 percent higher earnings than worker B (10 extra points on the occupational English requirements scale, valued at 1.3 percent higher earnings per point). Worker C gets this higher earnings for two reasons: first, he has a higher proficiency in English, and second, he moves to an occupation that is suited to his superior English skills. Thus, the 13 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.

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 characterises worker C is missing in the case of worker D. Hence, while worker D earns 3 percent more than worker B, which is the increase in earnings associated with the higher proficiency in English, he earns 10 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 13 percent increase in earnings for worker C compared to worker B, 3 percentage points appear to be due to the acquisition of the superior English skills, and 10 percentage points appear to be due to worker C having moved to an occupation where the superior English skills can be used effectively.

Table 3
Earnings and Skill Classification of Five Hypothetical Workers

	Required Level of English	Actual Level of English	Skill Classification	ln*Y*
A	30	30	Correct Match	10.03
B	40	40	Correct Match	10.16
C	50	50	Correct Match	10.29
D	40	50	Overqualified	10.19
E	40	30	Underqualified	10.11

Figure 3
Earnings of Five Hypothetical Workers Described in Table 3



The variable for an underqualified English score has a negative coefficient, of -0.005. This indicates that the person will incur a negative return of 0.5 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. Thus, consider a worker who has an actual English proficiency score of 30 who gains a position in an occupation that has an English requirement of 40, and compare this underqualified (in English) worker to workers in jobs with occupational English requirements of 30 and 40 who are correctly qualified (in English) for their jobs. Denote these workers as E, A and B, respectively. They are described in Table 3. Given the specification of the earnings equation adopted here, comparing workers B and E 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 0.013 per point, or 13 percent higher earnings). However, the fact that worker E is underqualified attracts an earnings penalty compared to a correctly qualified worker in this occupation (worker B). This earnings penalty is -0.005 per point on the English scale, and so it is 5

percent for worker E. Thus, the underqualified worker (worker E) in this comparison will earn 5 percent less than the correctly qualified worker B, who is employed in a job with the higher English requirement, but worker E earns 8 percent more (*i.e.*, 13 percent for the higher English requirement less 5 percent for being underqualified) than the correctly qualified worker A who is employed in a job with the lower English requirement. The earnings of these comparison workers are displayed in Figure 3.

The estimates of the earnings effects associated with being overqualified in terms of English skills, underqualified in terms of English skills, and for being correctly matched in terms of English skills, mirror the findings from the undereducation/overeducation literature. This indicates that the central ideas of this earlier literature generalize to other human capital skills. However, the findings presented above 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. In the first instance, 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 Realized Matches methodology. 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.[10]

[10] This check parallels concerns in the undereducation/overeducation literature – see, for example, Hartog (2000, p. 135). 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.

Table 4[(a)]
Estimates for English Requirements and Mismatch Variables for Alternative Scores of Actual English Proficiency, Foreign-Born Males Aged 25-64, 2000 US Census

Variable	Level			Importance		
	(i) [(b)]	(ii) [(c)]	(iii) [(d)]	(i) [(b)]	(ii) [(c)]	(iii) [(d)]
Level/Importance of English	0.013 (60.86)	0.014 (62.43)	0.013 (61.26)	0.010 (58.72)	0.011 (60.73)	0.010 (59.53)
Overqualified English	0.003 (21.86)	0.003 (17.79)	0.002 (19.03)	0.002 (17.08)	0.002 (12.23)	0.002 (15.43)
Underqualified English	-0.005 (12.18)	-0.008 (20.31)	-0.007 (16.07)	-0.006 (18.45)	-0.008 (27.82)	-0.007 (21.80)
Adjusted R^2	0.380	0.382	0.381	0.378	0.381	0.379

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

The results suggest that while the point estimate of the earnings effect associated with the underqualified English variable varies slightly with the choice of method for scoring workers' actual English proficiency, the general pattern of effects – in terms of sign, relative numerical magnitudes, and statistical significance, is not affected by the choice of algorithm for scaling a worker's proficiency in English.

B. Sensitivity to the Measure of Occupational English Requirements

The analyses reported above are based on occupational English requirements sourced from the O*NET database. As explained in Section II, this is a hybrid of the Job Analyst and Worker Self-Assessment methods for determining the level of English required in each occupation. The undereducation/overeducation literature has used three alternatives in this regard, with the third being a Realized Matches method. This uses the self-reported English proficiency of workers in each occupation to prepare the reference level

of English required to work 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.

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.[11]

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.[12] The full set of results from this analysis based on the Realized Matches procedure is presented in Appendix B.

[11] As the native born are almost all English-only speakers, and dominate the representation in the overwhelming majority of occupations, their inclusion in the sample used to form the benchmark English levels would drive these benchmarks towards 100. Moreover, the native born would be proficient in English regardless of their occupation.

[12] 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.

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

Variable	Model I[(a)]	Model II[(b)]	Model III[(c)]
English Requirement, Realized Matches Procedure	0.018 (79.18)	0.021 (79.56)	0.018 (80.08)
Overqualified English	0.002 (14.68)	0.002 (13.10)	0.002 (12.25)
Underqualified English	-0.005 (10.73)	-0.006 (15.54)	-0.006 (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 column (i) of Table 1 were used.

(b) When forming the occupational English requirements and workers' actual proficiency in English measures, the scores from column (ii) of Table 1 were used.

(c) When forming the occupational English requirements and workers' actual proficiency in English measures, the scores from column (iii) of Table 1 were used.

Source: 2000 US Census, 1% PUMS.

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 only focuses on the estimates labeled "Model I", which use the scores for the English proficiency categories given in column (i) of Table 1 to establish the reference level of English for each occupation.

The English requirement variable in the first of these alternative analyses, which gives the return to worker's correctly-matched English skills as required by the job, has a positive coefficient of 0.018, and a 't' value of almost 80. This compares with the coefficient of 0.013 and 't' value of 53.52 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 are associated with

each 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. Both effects are precisely estimated, with 't' values between 15 and 22. 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 in Table 5. This is the same value as that estimated using the O*NET Level of English information.

Hence, while the point estimates for the earnings effects differ slightly, the general pattern of effects for the required and mismatch English variables is the same across the various measures of occupational English requirements considered.[13]

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

Hartog (2000, p. 135) has noted that tests of the impact of non-linearity in the returns to required years of education show that the use of a quadratic functional form for the education variable does not impact on the conclusions of the undereducation/overeducation analysis. The extent to which this finding carries across to the analysis of occupational English language requirements is examined here.

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

[13] This mirrors the finding from the overeducation/undereducation literature. For example, Hartog (2000, p. 135) concludes that "These conclusions are not sensitive to the measure of required education…Sometimes estimated returns are affected by the type of measurement, but sometimes they are virtually identical…This conclusion is quite appealing in view of the measurement unreliabilities noted above."

Table 6
Estimates of Earnings Function With and Without English Overqualification and Underqualification Variables, with Quadratic Level/Importance Variables, Foreign-Born Males Aged 25-64, 2000 US Census

Variable	Level		Importance	
	(i)	(ii)	(i)	(ii)
Constant	5.327	5.398	5.394	5.501
	(100.83)	(100.91)	(102.10)	(102.94)
Log Weeks Worked	0.868	0.869	0.867	0.868
	(72.68)	(72.70)	(72.66)	(72.73)
Years of Education	0.037	0.036	0.038	0.038
	(43.93)	(43.43)	(46.86)	(45.98)
Experience	0.015	0.015	0.013	0.013
	(14.31)	(14.57)	(13.09)	(13.08)
Experience Squared/100	-0.024	-0.024	-0.021	-0.021
	(12.01)	(12.30)	(10.76)	(10.88)
Married	0.194	0.195	0.195	0.195
	(32.97)	(33.11)	(33.03)	(33.18)
South	-0.070	-0.070	-0.069	-0.070
	(11.68)	(11.75)	(11.55)	(11.65)
Metropolitan Area	0.108	0.116	0.106	0.114
	(4.04)	(4.33)	(3.96)	(4.28)
Veteran	-0.061	-0.058	-0.065	-0.061
	(4.72)	(4.49)	(4.97)	(4.67)
Blacks	-0.150	-0.149	-0.153	-0.152
	(14.55)	(14.41)	(14.78)	(14.64)
Years Since Migration (YSM)	0.010	0.011	0.010	0.010
	(12.25)	(12.64)	(11.84)	(12.41)
YSM Squared/100	-0.008	-0.009	-0.007	-0.008
	(4.10)	(4.46)	(3.80)	(4.31)
Level of English	0.013	0.007	(a)	(a)
	(60.86)	(9.45)		
Level of English Squared/100	(a)	0.007	(a)	(a)
		(7.22)		
Importance of English	(a)	(a)	0.010	0.003
			(58.72)	(5.81)
Importance of English Squared/100	(a)	(a)	(a)	0.008
				(11.85)
Overqualified English	0.003	0.003	0.002	0.002
	(21.86)	(22.42)	(17.08)	(17.75)
Underqualified English	-0.005	-0.005	-0.006	-0.006
	(12.18)	(11.74)	(18.45)	(17.90)
Adjusted R^2	0.380	0.380	0.378	0.379
Sample Size	84,172	84,172	84,172	84,172

Note: Heteroskedastic-consistent 't' statistics in parentheses. Source: 2000 US Census, 1% PUMS.

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 a little less than 1 percent at a level of English of 20, to around 2 percent at a level of English of 80 (compared to the uniform partial effect of 1.3 in the linear specification).[14] Similarly, for the importance of English, the partial effect changes from one-half of 1 percent at an importance score of 20, to 1.58 percent at an importance score of 80 (compared to the uniform effect of one percentage point in the linear specification). However, the coefficients on the overqualification (in English) and underqualification (in English) variables are the same as reported previously on the basis of the simpler specification. Moreover, the statistical significance of the mismatch variables is unaltered by the modification to the way the English language requirements variable is modeled. Hence, the conclusion of Hartog (2000, p. 135), that the education mismatch effects are not due to non-linearity in the returns to schooling, is mirrored clearly by the Table 6 results.

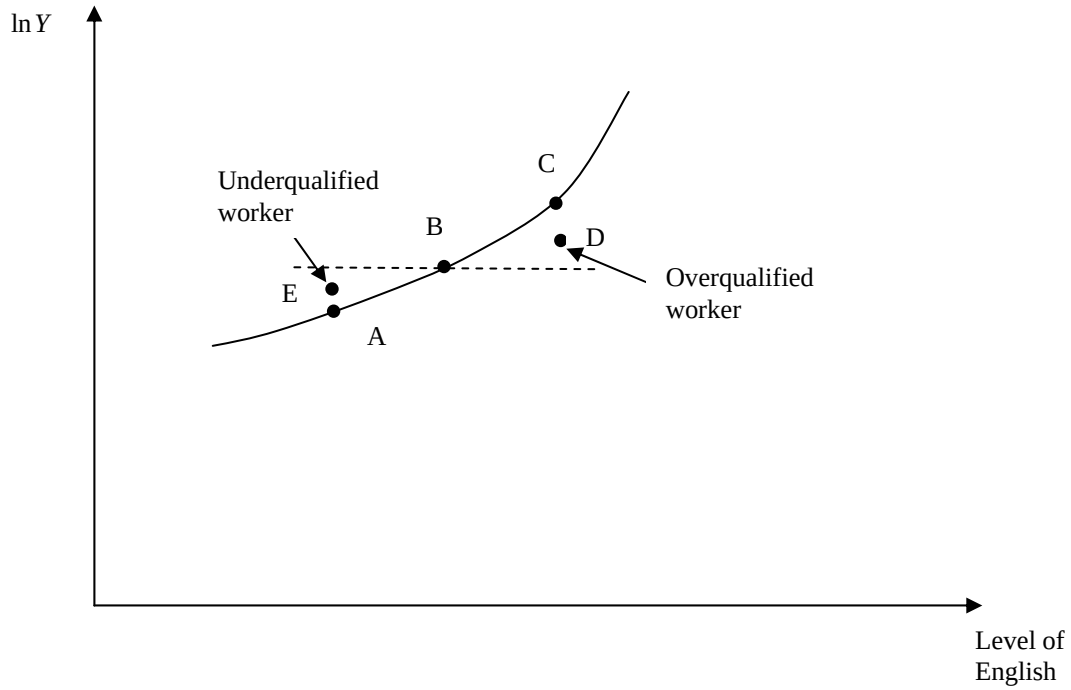
The use of a quadratic for the occupational English language requirements, however, has an effect on the dollar value of the earnings effects associated with overeducation and undereducation. These will, given the non-linearity in the English language requirements variable, vary across the scale of the variable, being larger for a level of English above 42, and smaller for levels of English below that. To illustrate these, the Table 3 calculations for five hypothetical workers were repeated, using the results of Table 6, column (ii). These calculations are presented in Table 7, and illustrated in Figure 4. The Table 7 predictions have been calibrated so that the predicted earnings for worker B are the same as in Table 3.

[14] Occupations with a level of English of 20 include earth drillers and tool and die makers. Those with a level of English of 80 include postsecondary teachers and lawyers. Occupations with an importance of English of 20 include glaziers and shoe machine operators and tenders, while those with an importance of English of 80 include first line supervisors/managers of correctional officers and respiratory therapists.

Table 7
Earnings and Skill Classification of Five Hypothetical Workers

	Required Level of English	Actual Level of English	Skill Classification	ln*Y*
A	30	30	Correct Match	10.08
B	40	40	Correct Match	10.16
C	50	50	Correct Match	10.25
D	40	50	Overqualified	10.19
E	40	30	Underqualified	10.11

Figure 4
Earnings of Five Hypothetical Workers Described in Table 3



The data in Table 7 show that the difference in earnings between an underqualified worker (worker A) and a correctly matched worker (worker E) is less than reported earlier. This pattern will arise because the calculations for the underqualified worker in Table 7 are undertaken over the Levels of English of 30-40, that is, at values less than the threshold of 42 at which the partial effect in the non-linear specification becomes greater than that of the linear specification. In comparison, the difference in earnings between an overqualified worker (worker D) and a correctly matched worker (worker C) is greater in

Table 7 than in Table 3. These differences, however, do not affect the interpretation of the general patterns in the findings reported in Table 2.

VI. SUMMARY AND CONCLUSION

This paper examines the application of the educational mismatch hypothesis to occupational English language requirements. The theoretical framework employed is grounded in the undereducation/overeducation literature. Empirical analyses throughout this paper have used data on foreign-born adult (age 25 to 64) men from the 2000 US Census (one percent microdata sample) and information on occupational English language requirements obtained from the O*NET database to investigate whether the propositions of the human capital skills/matching literature carry across to the study of language skills.

The measurement procedures established in the undereducation/overeducation literature allowed the identification of workers whose English language skills were correctly matched to their job requirements, those who were overqualified in terms of their English skills, and those who were underqualified in terms of their English skills. The earnings consequences are computed of the mismatch of a worker's English skills to the requirements (level and importance) of the occupation in which they work.

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 relatively minor payoff to the acquisition of English language skills within one's occupation. Second, there is a more important payoff to mobility to an occupation better suited to the superior English skills. Thus, overqualified workers earn modest rewards for their excess endowment of English skills, but these rewards are far less than if they moved to jobs in which they were correctly matched to the English requirements of their job. However, as underqualified workers gain the second of these payoffs but not the first, they have much higher earnings than their counterparts with relatively low English skills who work in jobs where their skills are

correctly matched. Underqualified workers do incur earnings penalties for their skill inadequacies, compared to workers in their occupation who have the right level of English skills for their job.

To ensure the results presented are robust, several checks were conducted to test their sensitivity 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 in the earnings equation. The tests confirmed the robustness of the estimates.

The results from this paper have demonstrated that the conceptual framework of the undereducation/overeducation literature is useful in the study of language skills and labor market outcomes. Future research could seek to establish whether these empirical regularities regarding language skills, occupation and earnings, for the US 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 signify sub-optimal and inefficient resources allocations, policy actions might encourage better job matching in the labor market. To date, effort appears to have been directed towards resolving issues of undereducation and overeducation. The results presented in this paper suggest that attention should also be paid to the issue of language skill mismatch. This could be an important role for immigrant settlement policies.

RERERENCES

Bleakley, Hoyt and Chin, Aimee, (2004). "Language Skills and Earnings: Evidence from Childhood Immigrants", *Review of Economics and Statistics*, Vol. 86, No. 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*, Vol.25, No. 2, pp.73-95.

Chiswick, Barry R. and Miller, Paul W., (2006a). "Occupational Language Requirements and the Value of English in the US Labor Market", mimeo, Department of Economics, The University of Illinois at Chicago.

Chiswick, Barry R. and Miller, Paul W., (2006b). "Why is the Payoff to Schooling Smaller for Immigrants?", mimeo, Department of Economics, The University of Illinois at Chicago.

Chiswick, Barry R. and Miller, Paul W., (2007). "Modelling Immigrants' Language Skills", *Research in Labor Economics*, (forthcoming).

Duncan, Greg and Hoffman, Saul D., (1981). "The Incidence and Wage Effects of Overeducation", *Economics of Education Review*, Vol. 1, No. 1, pp.75-86.

Espenshade, Thomas J. and Fu, Haishan, (1997). "An Analysis of English-Language Proficiency Among U.S. Immigrants", *American Sociological Review*, Vol. 62, No. 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*, Vol. 2, No. 3, pp.226-236.

Evans, Mariah D.R., (1987). "Language Skill, Language Usage and Opportunity: Immigrants in the Australian Labour Market", *Sociology*, Vol. 21, No. 2, pp.253-274.

Hartog, Joop, (2000). "Over-Education and Earnings: Where Are We, Where Should We Go?", Economics of Education Review, Vol. 19, No. 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*, Vol. 22, No. 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 from the O*NET Consortium described in Section II.

Definition of Population: Foreign-born men aged twenty-five to sixty-four. 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 binary 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 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): This is the overqualification variable measured in English scores, and it is derived as 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): This is the underqualification variable measured in English scores, and 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.

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

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

(d) When forming the occupational English requirements and workers' actual proficiency in English measures, the scores from column (i) of Table 1 were used.

(e) When forming the occupational English requirements and workers' actual proficiency in English measures, the scores from column (ii) of Table 1 were used.

(f) When forming the occupational English requirements and workers' actual proficiency in English measures, the scores from column (iii) of Table 1 were used.

Source: 2000 US Census, 1% PUMS.

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	0.263 (60.44)	0.317 (73.58)	0.328 (75.61)	0.321 (74.72)
Overqualified English	0.032 (6.22)	0.040 (7.83)	0.041 (8.29)	0.036 (7.06)
Underqualified English	-0.169 (30.34)	-0.109 (18.65)	-0.130 (20.95)	-0.122 (20.30)
$\overline{R}^2$	0.382	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, 1% PUMS.

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 0.263 to 0.328, while the effect of the overqualified English variable varies from 0.032 to 0.041. 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 -0.169. 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.[15]

[15] This issue does not appear to have been addressed in the overeducation/undereducation literature.