02 August 2007

Preliminary Draft
Not for quotation nor citation

MODELING IMMIGRANTS' LANGUAGE SKILLS*

Barry R. Chiswick
Department of Economics
University of Illinois at Chicago
and
IZA Institute for the Study of Labor

and

Paul W. Miller
Business School
University of Western Australia

Keywords: Immigrants, Language, Enclaves, Human Capital

JEL Codes: F22, J15, J24, J40

* This paper is for presentation at the "Conference on Immigration: Trends, Consequences and Prospects for the United States", September 9-10, University of Illinois at Chicago. 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.

02 August 2007

MODELING IMMIGRANTS' LANGUAGE SKILLS

ABSTRACT

One in nine people between the ages of 18 and 64 in the US, and every second foreign-born person in this age bracket, speaks Spanish at home. And whereas around 80 percent of adult immigrants in the US from non-English speaking countries other than Mexico are proficient in English, only about 50 percent of adult immigrants from Mexico are proficient. The use of a language other than English at home, and proficiency in English, are both analyzed in this paper using economic models and data from the 2000 US Census. The results demonstrate the importance of immigrants' educational attainment, their age at migration and years spent in the US to their language skills. The immigrants' mother tongue is also shown to affect their English proficiency; immigrants with a mother tongue more distant from English being less likely to be proficient. Finally, immigrants living in ethnic enclaves have lesser proficiency in English than immigrants who live in predominately English-speaking areas of the US. The results for females are generally very similar to those for males, the findings from an ordered probit approach to estimation are similar to the findings from a binary probit model, and the conclusions drawn from the analyses mirror those in studies based on the 1980 and 1990 US Censuses. Thus, the model of language skills presented appears to be remarkably robust across time and between the genders.

MODELING IMMIGRANTS' LANGUAGE SKILLS

I. INTRODUCTION

Immigrants typically fare quite poorly in the labor market in the US and in other immigrant receiving countries in the immediate post-arrival period. Using data from the 1990 US Census, Miller and Neo (2003) show that the earnings in 1989 of adult male recent arrivals in the US were up to 28 percent below the earnings of comparable native-born workers. This disadvantage arises because of the less-than-perfect international transferability of immigrants' human capital, their lack of knowledge of the institutions of the US labor market, and perhaps through discrimination. The earnings disadvantage is temporary for most groups: Immigrants tend to rapidly catch up with the native born, and much of the empirical research into immigrant labor market outcomes has focused on the factors that enhance this adjustment process (Chiswick (1978)(1979)). Attention has been directed at post-arrival investment in human capital in general, and destination-specific language capital in particular. Among immigrants in the US labor market, English language skills attract considerable wage premiums: Chiswick and Miller (1995) show that the earnings gain in 1989 associated with proficiency in English among adult male immigrants was approximately 17 percent, or the equivalent of around three years of schooling. This earnings increase is similar to that associated with dominant language proficiency in Canada, and is more than twice the increase in earnings associated with English language proficiency among immigrants in Australia. Destination language skills have been found to be important in non-English speaking destinations, such as Germany and Israel (Chiswick and Miller (1998)). Moreover, part of immigrants' economic progress captured by duration of residence variables may be linked to improvements in their English language skills (see, for example, McManus, Gould and Welch (1987)).

Presumably largely reflecting these rewards, the English language skills of the foreign born in the US improve rapidly with duration of residence. Among adult male workers who had lived in the US for fewer than four years, the 1990 US Census reveals that as few as 13 percent spoke English only. A further 44 percent spoke a language other than English at home, but spoke English very well. However, fully 43 percent of this group spoke a language other than English at home and reported that they spoke English "not

well" or "not at all" (Chiswick and Miller (1996)). Among those who had resided in the US for 21-30 years, however, monolingual English speakers were more prevalent (31 percent), and those with either limited or no English skills (*i.e.* they spoke English "not well" or "not at all") were far less prevalent (14 percent).

This paper examines the processes associated with the acquisition of English language skills among the foreign born in the US. It has a special focus on the role that ethnic networks and linguistic distance play in the acquisition of dominant language skills, and employs alternative measures of ethnic networks to that pioneered in Chiswick and Miller (1998)(2002)(2005a). Extensions of the analysis of language practice to consider origin-language retention are also presented. The paper also gives attention to females as well as to male immigrants, and consolidates a series of modifications to the language model made in various papers. Being based on data from the 2000 US Census, it permits an update of the evidence using the 1990 Census reported in Chiswick and Miller (1998)(2002)(2005a).

The structure of the paper is as follows. Section II outlines briefly the model of dominant language acquisition, introduced in Chiswick (1991) and Chiswick and Miller (1992), and subsequently developed in Chiswick and Miller (1995)(1998)(2005a). Section III outlines the data set to be used, the US 2000 Census of Population, and presents estimates of models of dominant language proficiency for males. Estimates from Binary Probit and Ordered Probit models are considered. Similar sets of estimates for females are presented in Section IV. Section V contains the analyses of origin-language retention among immigrants. These are also presented for both males and females. Concluding comments are provided in Section VI.

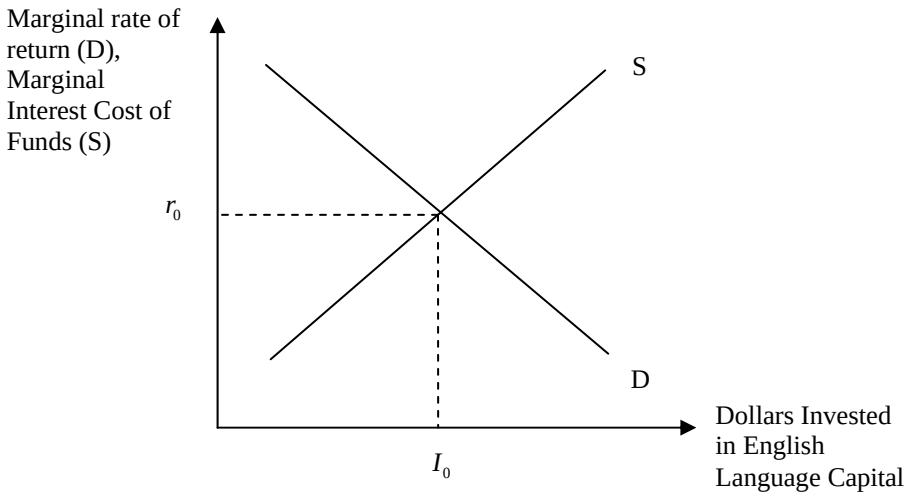
II. MODEL OF DOMINANT LANGUAGE PROFICIENCY

Immigrant decision making in relation to the learning of dominant language skills can be analyzed using a human capital framework. Destination language proficiency among immigrants is a form of human capital that is productive in consumption and/or labor market activities, it is costly in terms of time and out-of-pocket expenditures to acquire,

and it is embodied in the person. Thus, the optimal investment in destination language proficiency for immigrants is determined in the same manner as for other human capital investments: as the level of investment that will equate the marginal rate of return on the investment to the marginal interest cost of the funds they invest. This decision-making process is outlined in Figure 1 (which is based on Becker and Chiswick (1966)) for immigrants from a non-English speaking country.

The demand for language skills (D) is given by the marginal rate of return on the investment in these skills. The position of this curve depends on the costs of, and benefits from, the investment in language skills. The demand (or marginal return schedule) will be higher the lower are the costs and the greater are the benefits from investment in language capital. The costs of the investment include the direct costs of language classes, as well as the indirect costs of foregone earnings. The benefits from destination language acquisition include higher wages, lower chances of being unemployed, greater efficiency in consumption, and greater participation in social activities and political processes.

Figure 1
Supply and Demand for Funds for Investment by Immigrants From non-English speaking Countries in English Language Capital



The demand curve is downward sloping for several reasons. First, individuals will invest first in those dimensions of language skills that have the highest payoff (marginal returns)

in labor market and non-labor market activities. These are followed by less productive investments. Second, as previous investments raise the opportunity cost of time, the marginal rate of return declines even if the dollar value of the benefit is unchanged. Third, since investments take place over time, the greater the investments already made, other things the same, the fewer the time periods remaining in the future and hence the lower the return on additional investments. And, finally, as with most activities, whether investment, production or consumption, beyond some point diminishing returns set in the greater the intensity of the activity. The marginal product of an hour of language learning per day will, at some point, start declining the greater the number of hours in language study.

The supply of funds for investment in language skills (S) is given by the marginal interest cost of funds. This curve will be upward sloping because immigrants will use lower-cost sources of finance (own savings, family and friends) before they access more expensive sources. For these reasons, this curve will also be lower for those with greater access to resources, including greater wealth, for financing the investment in language capital.

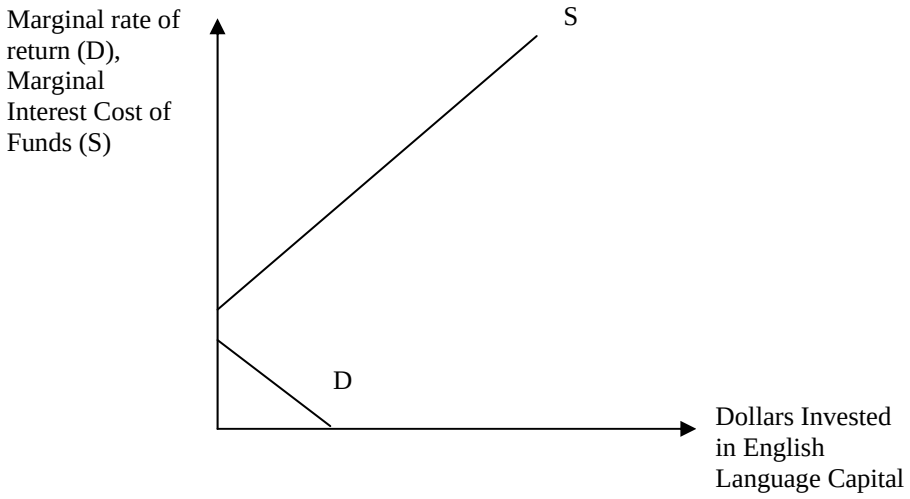
The intersection of the demand and supply curves gives an optimal level of English language proficiency (I_0) for the immigrant.

The estimating equation to be developed below can be thought of as a reduced form equation incorporating both the supply and demand conditions for funds for investment in language capital. The actual dollar amounts invested cannot be estimated, but other variables being the same, the immigrants' level of English language proficiency can serve as a proxy for the dollars invested. The explanatory variables considered below (*e.g.*, age at migration, educational attainment, years of residence in the US) may, for example, shift the demand curve outward or inward, resulting in higher or lower proficiency in English.[1]

[1] For example, due to greater efficiency in language acquisition, an immigrant who arrived in the US at 10 years of age would be expected to develop greater proficiency in English from any dollar investment than one who arrived in the US at 50 years of age. Thus, age at migration would be a shift variable in the demand for funds equation and affect investment levels and proficiency.

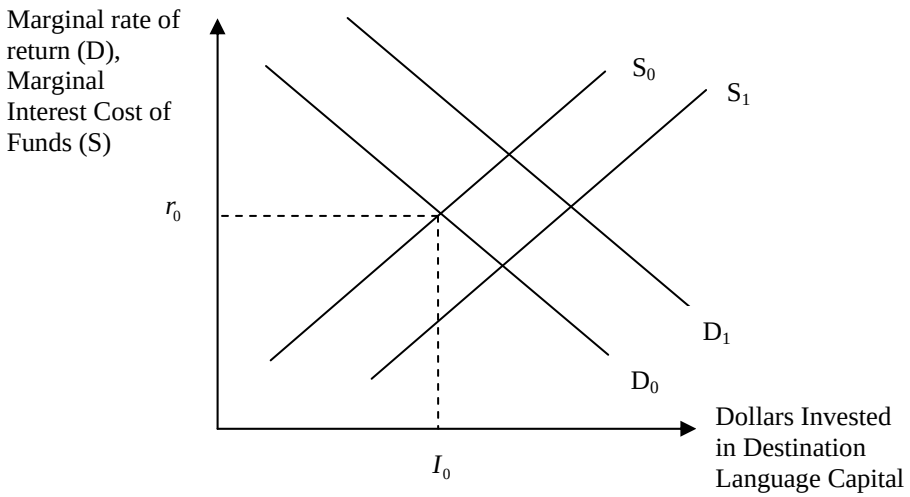
For an immigrant from an English-speaking country, however, the marginal rate of return on investments in English language training are so low that no or minimal investments would be made. This is illustrated in Figure 2, where there are no investments in post-migration English language training. Given the trivial magnitude of the investments by this group, models of English language acquisition have generally been applied only to non-English-speaking background immigrants.

Figure 2
No Investment in English Language Capital



The demand and supply curves will shift for a variety of reasons. The demand curve will shift to the right if an immigrant is more efficient in the production of language capital (D_1 compared to D_0 in Figure 3). This might be associated with a higher educational attainment, or being of a younger age at the time of migration. The supply curve for an immigrant with greater wealth will be to the right of that for an immigrant with relatively little wealth (S_1 compared to S_0 in Figure 3).

Figure 3
Shifts in Supply and Demand for Funds for Investment by Immigrants From non-English speaking Countries in English Language Capital



Variations in English proficiency among the immigrant population will therefore be linked to factors that shift the demand and supply curves in Figure 3. Chiswick and Miller (1992)(1995)(1998)(2005) categorise these as exposure, efficiency, and economic incentive factors. Hence, destination language proficiency (LANG) can be expressed as:

$$LANG = f(\text{Exposure, Efficiency, Economic Incentives}) \quad (1)$$

A. Language Measures

Three measures of language practice are used in the statistical analyses. The first is a binary measure (LANG2), which is set equal to one for individuals who speak only English at home, or if a language other than English is spoken in the home, the individual speaks English either "very well" or "well." The variable is set to zero where a language other than English is spoken in the home and the respondent speaks English either "not well" or "not at all." The second measure (LANG5) is also a measure of proficiency, and it is a polychotomous (five categories) variable, defined to include all proficiency categories contained in the relevant Census variables: (i) speaks English only at home; speaks a Language other than English at home and speaks English (ii) very well; (iii) well; (iv) not well; (v) not at all. The third language variable (MT) is a binary variable that records whether the individual speaks a language other than English at home (MT =

1) versus speaking only English at home (MT = 0). Where a language other than English is spoken at home it is assumed, for ease of discussion, to be the individual's mother tongue. This is a measure of origin language retention.

B. Exposure Factors

Exposure to the destination language can occur before or after immigration. The degree of pre-immigration exposure depends, in large part, on the extent to which English is used in the origin country. This could be due to a British/US colonial past, or a major US military presence, though with advances in telecommunications and the world-wide spread of American and British media and movies, and more recently the internet, it would seem that pre-immigration exposure could become reasonably widespread. A dichotomous variable for whether the origin was a colony of the United States or the United Kingdom (COLONY) is used to capture some of these influences..

Post-immigration experience will depend on two main factors. First, there is the time units of exposure to English. This is the extensive margin, and can be measured by the number of years since the immigrant came to the US to stay. A quadratic specification (YSM and YSMSQ) is used to allow the effect of an extra year in the United States to be larger in the early years than in subsequent years.[2] The variable assumes that the immigrant has lived continuously in the US from the time of arrival. Some immigrants, however, spend time outside the US after the initial migration. The potential effects of this sojourner migration, or to- and fro-migration, on English language skills can be assessed for those immigrants who arrived before 1995 by a dichotomous variable (ABROAD5) which is unity where the immigrant lived abroad five years ago, and is zero otherwise.

[2] While improvements in English skills can be expected with practice, most investment in language skills should occur just after migration, since investments in language skills tend to be more profitable if the period over which the benefits will be received is longer, since the opportunity cost of investment in language training is lower in the early period when wages are lower, and since the returns are greater if investments with high rates of return are made sooner rather than later. The complementarity in the labor market of language skills and schooling and post-migration labor market experience also encourages earlier investments. Consequently, a quadratic specification for duration of residence should be used.

Second, there is the intensive margin, which is the intensity of exposure per unit of time. The intensity of exposure per unit of time depends on the immigrant's neighborhood and family experiences. Ethnic enclaves can be expected to play a major role in the immigrant's language skills. An immigrant who lacks English language skills can avoid having to learn English by living in an area in which many others use his or her origin language. Similarly, working in a linguistic enclave can limit the benefits from acquiring English skills. These effects can be measured by the proportion of the population of the area, regardless of nativity, that speaks the immigrant's origin language (CONC). The top 25 non-English languages spoken at home are utilized in the construction of the CONC variables. These cover about 91 percent of the sample of immigrants from non-English speaking countries used in the analyses reported below.[3] Language rather than birthplace or ancestry is used as the defining element of an enclave on the grounds that it better measures the cultural linguistic concept developed in Chiswick and Miller (2005b). For example, the use of birthplace in the construction of this measure would encounter difficulties with bilingual (*e.g.*, Belgium) and multilingual countries (*e.g.*, India), and areas over which there is a common language used in many countries (*e.g.*, Spanish in Mexico, Central America and much of South America).

Of special interest in the work reported below is the most appropriate definition of "area" when attempting to capture these ethnic enclave effects in a model of destination language proficiency. Three alternatives are employed, and these are distinguished by the level of geographic identifiers used in their construction. The first is based on the State (50 States and the District of Columbia) of residence (CONC-State, which was used in earlier Chiswick and Miller research). The second is based on the Super Public Use Microdata Areas (Super-PUMAs) used in the 1 percent Public Use Microdata Sample (PUMS) of the 2000 US Census (CONC-Super).[4] All 532 separate areas are utilised for the construction of this variable. The third variable is constructed using the information

[3] See Appendix A for a list of these languages and their shares of the immigrant population.

[4] A Super-PUMA is a geographic entity that comprises at least 400,000 people.

on Metropolitan Statistical Areas (CONC-MSA).[5] The 106 separate regions are used in forming this measure using the 1 percent sample. In this instance, immigrants living outside metropolitan areas are assigned the value of the concentration measure constructed for the non-metropolitan components of their state of residence. The CONC variable is set equal to zero in all three definitions for those reporting a language that is not among the top 25 languages on the grounds that the density or concentration of speakers of these languages is too low to matter.

The CONC variables constructed may not capture the intended influences outside metropolitan areas. This is for several reasons, depending on the particular measure employed. First, where the CONC variable is computed at the state level, it would be expected that it would over-estimate the minority language concentration in non-metropolitan areas. Second, where the CONC variable is computed from the more disaggregated data on MSA-PMSAs, there is insufficient detail on the Census files to permit identification of non-metropolitan areas in each of the states. Accordingly, a single non-metropolitan dichotomous variable (NON-MET) is included in the estimation equation. A variable for the southern states (SOUTH) is also included to capture regional influences. An alternative specification (described below) of regional variables is used when the regression analysis is limited to those born in Mexico.

Language practice within the family will also influence the individual's dominant language proficiency. Chiswick, Lee and Miller (2005a)(2005b) show that, due to similarities in the observed and unobserved characteristics of family members, there are links between their dominant language fluency. These factors include assortative mating, genetic and home investment linkages between parents and children, language learning in the family, migration as a family unit and systematic reporting errors within a household, as well as the similarity in the processes governing dominant language proficiency for family members. The correlations in language proficiency are stronger for spouses than

[5] A metropolitan area is one of a large population nucleus, together with adjacent communities that have a high degree of economic and social integration with that nucleus. Where a metropolitan area has 1 million people or more, two or more primary metropolitan statistical areas (PMSAs) may be defined within it. Information on the Metropolitan Statistical Areas (MSA) and Primary Metropolitan Statistical Areas contained in the variable MSA-PMSA1 is used in the construction of the CONC-MSA variable.

for parent-offspring combinations. The similarities in the underlying factors that give rise to these outcomes would be expected to be more apparent for those married prior to immigration (where marriage is more likely to be to a spouse from the same country of birth with the same linguistic background) than for those married after immigration. Where marriage takes place after immigration, it is more likely to be to a person who is not proficient in the immigrant's mother tongue. This may encourage the use of the dominant language. Accordingly, a marital status variable (MARRIED is unity if married, spouse present) is also employed in the model. It is not possible in the 2000 Census data to distinguish between pre- and post-migration marriages.[6]

The presence of children in the household could have a range of effects on immigrants' dominant language proficiency. Four channels have been identified in the literature. The first concerns children as teachers, based on the greater ability of children to learn new languages, and the intense exposure to the destination language in schools (Long (1990), Newport (1990), Service and Clark (1993)).[7] The rapid learning of the dominant language among children enables them to assist the development of the dominant language skills of their parents.

The second is children as translators. As children learn the dominant language, they move into a position where they can serve as translators for their parents. This possibility lessens the need for parents to acquire dominant language skills, at least from the perspective of consumption and home production. Children are unlikely to be able to serve this function in the workplace, other than possibility having a role in the context of the self-employed. This effect is likely to be stronger for mothers than for fathers.

Third, children can affect labor supply, particularly among females. To the extent that investments in language skills are made in anticipation of labor market activity, and to

[6] Year of first marriage was last asked in the 1980 Census.

[7] However, Fathman's (1975) review of the evidence indicated that "younger children are not necessarily better second language learners than adults in all respects", with her research showing that "after puberty the ability to learn rules, to make generalizations or to memorize patterns may be more fully developed".

the extent that the workplace provides an environment conducive to the further development of dominant language skills, reduced labor supply can dampen dominant language proficiency. This effect is also likely to be stronger for mothers than for fathers.

Fourth, where parents seek to transmit the culture of their country of origin, they may encourage the learning of their origin language among their children. Origin-language use within the home may therefore compete with dominant language use, with the potential to limit the development of dominant language skills among all family members.

Thus, while the learning of dominant language skills from their children will have a positive influence on the dominant language skills of their parents, the remaining three factors will tend to dampen the incentives for parents to acquire dominant language skills. As such, the sign of the overall effect of children on parental language skills is ambiguous. The effects of children on parental language skills would be expected to differ between the mother and the father, being less positive or more negative for the mother than for the father.

C. Efficiency Factors

There are four important measurable efficiency factors that can influence the development of dominant language skills among immigrants: age at migration, educational attainment, refugee status, and linguistic distance.

The young appear, for biological reasons, to have a greater capacity to learn a new language than do older individuals (Long (1990), Newport (1990), Service and Clark (1993)).[8] Age at migration (AGE) would therefore be expected to have a negative impact on dominant language skills. The effect of age at migration is measured by the partial effect of age (AGE) when years since migration are held constant. The age variable is entered in quadratic form, Age and Age Squared, in the analysis reported in the text.

[8] Although, see Fathman (1975) for an alternative perspective.

Appendix B contrasts this measure with age at migration entered as a set of dichotomous variables to test for discontinuities in the effects of age at migration.

Similarly, educational attainment (EDUC) is expected to be closely related to dominant language outcomes. The better educated may have technically superior language production functions. This could arise through the better educated having greater innate learning ability or unmeasured variables that enhance both forms of human capital. Or it could be that having greater knowledge of one's own language enhances the ability to learn other languages. It is also likely that the destination language, particularly where it is an international language such as English, may have been learned as part of the curriculum in either secondary school or tertiary studies abroad.[9]

The difficulty in learning a second language depends in part on the person's mother tongue. The argument here can be put as follows: it should be more difficult for a Chinese speaker to learn French than it is for a Spanish speaker to learn French because the differences between the languages are that much greater in the former case than in the latter case. In other words, the "linguistic distance" between Chinese and French is greater than the distance between Spanish and French. The greater the linguistic distance between the destination and origin language, the lower would be the efficiency of an immigrant for learning the destination language.

This concept of linguistic distance has been developed by Chiswick and Miller (1998, 2005b). Their measure is based on the ability of Americans to learn a variety of languages in fixed periods of time. The lower the scores on a standardized proficiency test, the greater the assumed distance between these languages and English. The equivalences outlined in Table 1 of Chiswick and Miller (2005b) are used in the data analyses reported below. In the case of those who report that they speak only English at home, the mean of the linguistic distance scores of immigrants in the US from the person's country of origin was used. Fully 97 percent of the sample have valid data for

[9] This last factor could not explain the greater level of Hebrew language proficiency of immigrants with a higher level of secular schooling in Israel (Chiswick and Repetto (2001)).

this measure. The remaining individuals are assigned the sample mean (their exclusion from the analyses yields similar results).

Linguistic distance may also be related to the degree of self-selection in immigration. Individuals with a mother tongue more distant from English, perceiving greater difficulty learning English, may only migrate if they have relatively high levels of unobservables that are related to the ability to learn English, and with immigrant adjustment in general.

Admission criteria may be relevant for understanding immigrant adjustment (see Chiswick, Lee and Miller (2005c)). Unfortunately, the US Census does not provide information on the visa used at entry, or the current visa status, other than whether the immigrant has become a naturalized citizen. Yet, research suggests that refugees experience a different adjustment than family or economic immigrants. Refugee status may impact on dominant language skills because refugees tend to be less favorably selected for a successful adjustment in the destination than are economic migrants. The less-intense selection arises because of the greater importance of factors in the migration decision other than the expectation of economic success. Refugees often have less time to prepare for the move. The refugee variable (REFUGEE) is based on country of birth, period of immigration, and age at migration. The latter criterion permits refugee status to influence dominant language outcomes only where the person entered the US as an adult.[10] A variable for US citizenship is not included since a degree of proficiency in English is generally required to become a naturalized citizen.

D. Economic Factors

Economic incentives for dominant language proficiency are central to the model outlined above. However, finding empirical counterparts to this set of factors is difficult. Only variables that broadly correspond to the underlying influences can be considered. Hence, from the theoretical perspective, it is desirable to include the expected increments in earnings for each individual in the empirical applications. While this is not possible, it is

[10] As an example, adult immigrants from Cuba would be classified as refugees for this analysis if they entered the US after Castro came to power in 1959, but not earlier Cuban immigrants, or those who came to US as children.

known that there are strong links between educational attainment and the economic returns from becoming proficient in the dominant language, and this suggests that the individual's level of education (EDUC) may serve as a proxy for the expected economic returns for the investment in dominant language skills.

The incentive for an immigrant to acquire English proficiency will be greater the longer the expected duration in the US, as this will be associated with greater returns from any given investment. It is expected that the degree of return migration and the degree of favorable self-selection in immigration will vary with the distance of the origin country from the US. Greater geographic distance implies a higher cost of migration and of return migration. This should deter the less able and be associated with better dominant language skills among those who do immigrate. It also implies a lower propensity for return migration which should also be associated with greater proficiency in English. This distance effect is captured through a variable for the number of thousands of miles (MILES) from the major city in the origin country to New York, Miami or Los Angeles, whichever is the shorter. A quadratic specification is used.

When the analysis is limited to immigrants from Mexico, the geographic distance variable is computed with reference to the capital of their state of residence, and three cites in Mexico, namely Mexico City, Tijuana and Ciudaf Juarez. Two alternative continuous measures were considered, namely the distance between Mexico City and the capital of the immigrant's current state of residence, and the minimum of the direct line distance between the capital of their current state of residence and either Tijuana or Ciudaf Juarez. The latter measure is used in the statistical analyses reported below as it yielded slightly stronger results. Moreover, in the analysis limited to immigrants from Mexico, an alternative measure of the regional variable for the US is employed. Further comment on these is provided below.

Hence, the empirical counterpart to equation (1) that is the basis for the analysis that follows is:

LANG = f(Educational Attainment, Age at Migration, Age at Migration Squared, YSM, YSMSQ, ABROAD5, MARRIED, Children, NON-MET, SOUTH, MILES, MILESQ, Linguistic Distance, CONC, COLONY, REFUGEE)

(2)

Following Chiswick and Miller (2001), the estimating equation includes five variables based partly on country of birth, namely the proportion of individuals living in the same region as the immigrant that speak his home language (CONC), whether the person is a refugee (REFUGEE), whether the origin is a former British or American colony (COLONY), linguistic distance, and miles of the country of origin from the US.[11] Unlike dichotomous variables for country of birth, these variables have behavioral interpretations, and they provide for greater understanding of the factors affecting language practice among immigrants. Chiswick and Miller (2001) show that the behavioral variables based on birthplace provide almost as much explanatory power as the birthplace dummy variables. Accordingly, birthplace fixed effects are not included in the model.

The data for the estimations presented below are from the 2000 Census of Population, Public Use Microdata Sample, and are for the 1 percent sample of the foreign-born adult (25-64 year old) men and women from non-English speaking countries.[12] This age bracket is the group of immigrants for whom the issues surrounding language choice are most acute. Separate analyses are conducted for men and women, and the extent to which this is necessary is examined. The analyses are performed overall and separately for immigrants from Mexico and all other countries. Mexico is the largest single source country, providing over one-third of the men and women in the sample. Moreover, Mexican migrants have much lower levels of skills (among adult males they have 8 years of schooling compared to 13 years for other immigrants) and a much greater proportion

[11] The emigration rate variable employed by Chiswick and Miller (2001) is not used here, as the information is dated and more recent comparable data do not appear to be available. These data were also affected by the presence of illegal immigrants in 1980 and their receiving amnesty by 1990.

[12] Immigrants from the main English-speaking countries (UK, Ireland, Canada, Australia, New Zealand and the Caribbean) are excluded as, for the reasons given in Section II, the language issues do not exist to any great extent for this group.

of illegal aliens than migrants from other countries, and they may be of special interest for these reasons. Moreover, Mexico and Canada are the only countries sharing a land border with the United States. The variables are defined in detail, and the means and standard deviations are reported, in Appendix A.

III. ESTIMATES: MALE IMMIGRANTS

Table 1 reports the basic regression analyses for foreign born men. The first column of this table lists estimates of a probit model examining variations in the summary measure of English proficiency given by the binary variable LANG2 described above. Three figures are reported in each cell for the binary probit models. The first is the estimated coefficient for the probit index; the second is the associated 't' statistic; and the third the marginal effect of the variable on the probability of being proficient in English. As there are multiple marginal effects with the ordered probit model (one for each of the language categories), only the estimated coefficient for the probit index and the 't' statistic are presented.

These estimates largely accord with the results reported in previous studies using the 1990 Census (see, for example, Chiswick and Miller (2005a)). English proficiency increases with years of schooling, with the partial effect of an additional year of education being an improvement of 3.2 percentage points in the predicted probability of being proficient in English.[13] In comparison, English proficiency decreases at an increasing rate with age at migration. That is, the older an immigrant is at the time of entry into the US the less likely he is to become proficient in English, and this effect gets stronger with age. This is a reflection of the phenomenon established in the linguistics literature of language skills being more difficult to acquire for older than for younger individuals. Separate analyses (not presented here) indicate that the pattern of age effects established in Table 1 carries across to an alternative specification of the estimating

[13] The partial effects have been computed using the formula $\phi(X\beta)\beta_k$ for continuous variables (where ϕ is the standard normal density function), and as differences in predictions for groups distinguished within the dichotomous variables. For example, the partial effect for marital status (MS) is the difference between the predicted rate of proficiency for those who are married and that for those who are not married, where these predictions are sample averages.

equation based on a number of dichotomous variables for age at migration. These analyses suggest that age at migration does not matter among immigrants from countries other than Mexico up to around age 15. Increases in age at migration are associated with lower rates of English proficiency beyond this threshold. For immigrants from Mexico, however, increases in age at migration are associated with lower rates of English proficiency across the full range of ages at migration represented in the data. That is, for this birthplace group there does not appear to be any critical age for the learning of English as a second language in the US.

Table 1
Probit Estimates of Language Models, Adult Foreign Born Men by Origin, 2000

Variables	Total Sample		Immigrants from All Countries except Mexico	Immigrants from Mexico
	Probit	Ordered Probit	Probit	Probit
Constant	-0.584 (6.31)	0.882 (13.43)	0.293 (2.21)	-0.920 (6.07)
Education	0.107 (88.89) [0.032]	0.084 (95.84)	0.122 (68.18) [0.024]	0.082 (47.48) [0.033]
Age at Migration	-0.008 (1.84) [-0.010]	-0.008 (2.62)	-0.044 (7.16) [-0.010]	-0.001 (0.09) [-0.009]
Age at Migration Squared/100	-0.025 (4.83) [-0.010]	-0.021 (5.94)	0.009 (1.30) [-0.010]	-0.028 (3.36) [-0.009]
Years Since Migration (YSM)	0.073 (43.83) [0.020]	0.053 (48.54)	0.080 (34.92) [0.017]	0.071 (28.00) [0.022]
YSM Squared/100	-0.052 (12.09) [0.020]	-0.017 (6.46)	-0.063 (10.75) [0.017]	-0.054 (8.46) [0.022]
Abroad 5 years ago	-0.358 (10.80) [-0.121]	-0.242 (9.80)	-0.415 (8.65) [-0.102]	-0.293 (6.43) [-0.116]
Married	0.143 (11.61) [0.044]	0.074 (8.46)	0.120 (6.81) [0.024]	0.172 (9.90) [0.069]
With own children under 6 years only	-0.055 (3.09) [-0.017]	-0.082 (6.57)	-0.024 (0.91) [-0.005]	-0.037 (1.45) [-0.015]
With own children 6 to 17 years only	-0.066 (4.60) [-0.020]	-0.076 (7.62)	-0.062 (3.15) [-0.013]	-0.027 (1.27) [-0.011]

With own children under 6 years and 6 to 17 years	-0.093 (5.55) [-0.029]	-0.112 (9.27)	-0.022 (0.85) [-0.005]	-0.074 (3.30) [-0.030]
Non Metropolitan	0.053 (1.09) [0.016]	0.024 (0.67)	0.301 (2.63) [0.050]	0.090 (1.68) [0.036]
South	0.072 (6.04) [0.021]	0.068 (8.15)	0.103 (5.89) [0.020]	0.002 (0.14) [0.001]
Miles ('000) From Origin	0.258 (19.11) [0.029]	0.243 (26.82)	0.227 (13.73) [0.016]	-0.153 (1.25) [-0.030]
Miles ('000) From Origin Squared	-0.022 (14.85) [0.029]	-0.027 (27.62)	-0.021 (12.62) [0.016]	0.038 (0.74) [-0.030]
Linguistic Distance	-1.305 (25.03) [-0.395]	-0.943 (26.61)	-1.396 (25.88) [-0.278]	(a)
Minority Language Concentration CONC-STATE	-0.014 (18.75) [-0.004]	-0.016 (29.45)	-0.013 (9.42) [-0.003]	-0.010 (6.24) [-0.004]
Colony	0.800 (30.21) [0.189]	0.591 (42.69)	0.797 (28.81) [0.123]	(a)
Refugee	-0.236 (9.31) [-0.077]	-0.072 (3.85)	-0.236 (8.87) [-0.053]	(a)
μ_1	(a)	1.082 (224.10)	(a)	(a)
μ_2	(a)	1.982 (410.08)	(a)	(a)
μ_3	(a)	3.369 (475.70)	(a)	(a)
Chi-Squared	30226.76	37258.38	16965.45	6217.43
Prediction Success Rate	77.95	46.17	84.10	67.99
Sample Size	85865	85865	54001	31864

Source: US Census of Population, 2000, Public Use Microdata Sample, 1 Percent Sample.
Note: Figures in parentheses are 't' statistics, and the figures in square brackets for the binary probit models are partial effects, with effects for variables entered into the model in quadratic form being evaluated at the mean and listed for both terms of the quadratic; (a) Variable not relevant.

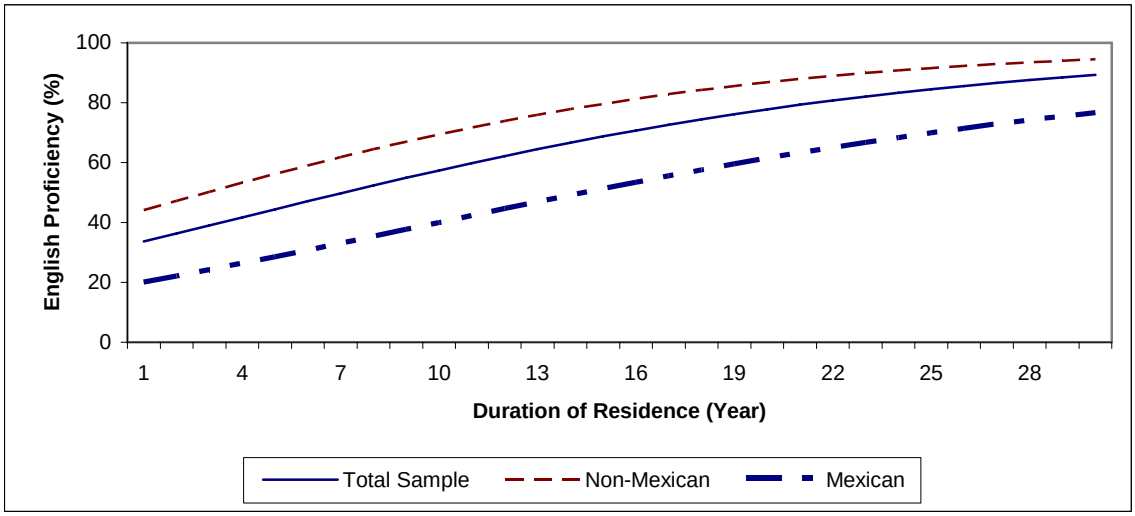
The immigrant adjustment process summarized in the years since migration variable is a strong influence on English proficiency. According to the estimates, English proficiency

improves at a decreasing rate with duration of residence in the US.[14] This improvement is registered across all levels of residency in the US recorded in the sample. In the first few years after arrival, the improvement in English proficiency is about 2.5 percentage points per year. At 10 years of residence, English proficiency improves by close to 2 percentage points per year. Even at 20 years there is improvement in the rate of English proficiency of around 1.8 percentage points per year of residence in the US. The relationship between English language proficiency and years in the US is portrayed in Figure 4. The profiles in this figure have been calibrated so that each one passes through the mean rate of English proficiency for the particular group when that group's duration of residence is equal to the group-specific mean.

Analyses performed for the 1980 Census and the 1990 Census (Chiswick and Miller 1992, 1998) show a similar pattern of a steeper rise in proficiency rates in the early years after immigration, with the rate of increase diminishing with duration. Selected partial effects of years since migration on English proficiency for male immigrants aged 25-64 years from non-English speaking countries and for those from Mexico are listed in Table 2. Across each year of data, the effect of years since migration is stronger for immigrants from Mexico. The effects of years since migration also get stronger across cohorts. That the increase in proficiency with duration repeats itself in repeated censuses suggests that it is reflecting a longitudinal phenomenon, rather than merely a decline in the linguistic proficiency of more recent cohorts or selective emigration of the least proficient immigrants in each arrival cohort.

[14] As Age = (Age at Migration + YSM), the model is $LANG = \beta_0 + \beta_1(Age\ at\ Mig. + YSM) + \beta_2 YSM + \ldots$, and so $\partial LANG / \partial YSM = \beta_1 + \beta_2$, or the effect of one year in the US plus the effect of an extra year of age. As the effects associated with years in the US are far stronger than those associated with age, the same pattern as established here carries over to the interpretation based on both β_1 and β_2, though as $\beta_1 < 0$ the years since migration effects are smaller. The discussion here has a focus on β_2.

Figure 4
Predicted English Proficiency by Duration of Residence in the US, Adult Foreign Born Men by Origin, 2000 US Census



Source: Authors' calculations based on Table 1.

Table 2
Partial Effect of Years since Migration on English Proficiency, Adult Foreign Born Men by Origin, 1980 1990 and 2000

	1980		1990		2000	
Years Since Migration	**Non-English Speaking Countries**	**Mexico**	**Non-English Speaking Countries**	**Mexico**	**Non-English Speaking Countries**	**Mexico**
5	1.62	2.15	1.95	2.50	2.41	2.61
10	1.34	1.80	1.70	2.20	2.23	2.39
15	1.06	1.45	1.45	1.90	2.05	2.18
20	0.78	1.10	1.20	1.60	1.86	1.96
25	0.50	0.75	0.95	1.30	1.68	1.74

Source: 1980: Chiswick and Miller (1992); 1990: Chiswick and Miller (1998); 2000; This paper, Table 1.

Just as the length of time an immigrant has spent in the US has a pronounced positive impact on English proficiency, spending time abroad after immigration diminishes English proficiency. Hence immigrants who came to the US to stay more than five years ago but who lived abroad in 1995 have a rate of English proficiency around 12 percentage points less than immigrants who were living in the US five years ago. This impact is the equivalent of the improvement that comes about through the first four years of residence in the US (see Figure 4). The intermittent nature of the stay among

sojourners, and perhaps the expectation of a relatively short future stay in the US among them, should be viewed as a major negative influence on immigrants' English language skills.

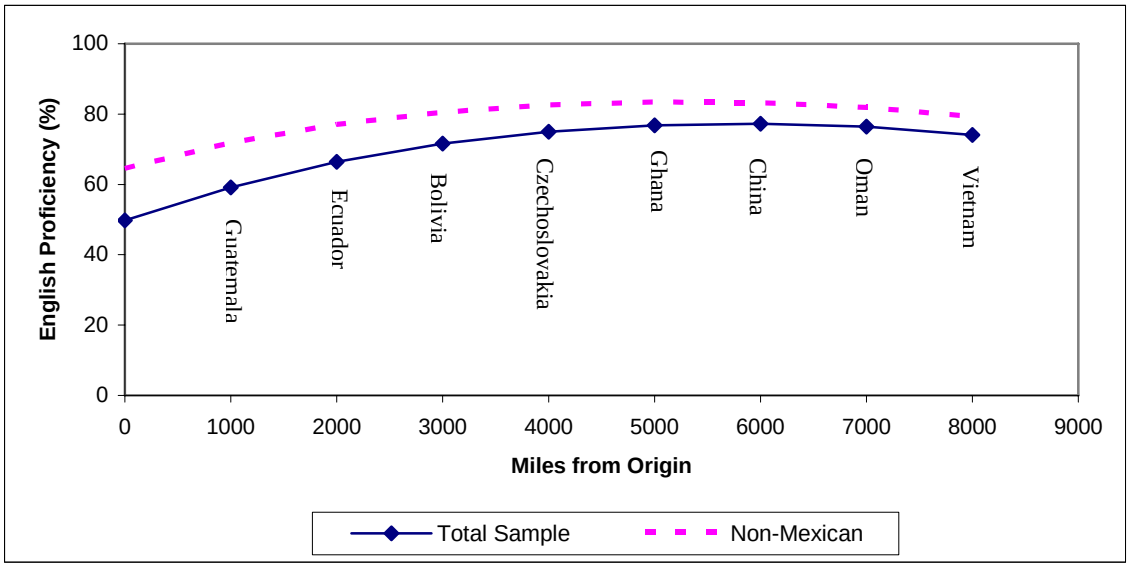
Immigrants who are married have a rate of English proficiency that is 4 percentage points above that of their non-married counterparts. As noted above, while it is desirable when modeling dominant language proficiency to be able to distinguish between marriage prior to migration (expected negative influence on dominant language skills) and marriage after migration (expected positive influence on dominant language skills where the marriage is to a dominant language speaker), this cannot be done with the 2000 Census data. The positive coefficient of the marriage variable on the rate of dominant language proficiency shows that the positive factors dominate.

Children are associated with a slight reduction in the English language skills of their fathers. Where the family only has a child (or children) under 6 years of age, or where it only has a child (or children) 6-17 years of age, the rate of English proficiency is reduced by 1.7 to 2.0 percentage points compared to having no children. This is the same magnitude of impact that would be associated with a reduction in educational attainment of around two-thirds of one year. In each of these cases there may be one or more children in the family, though only children cannot be distinguished from multiple child cases. However, where the family has children under 6 years and between 6 and 17 years (that is, there are at least two children present) the English proficiency of the adult males is reduced by 2.9 percentage points, an effect that is equivalent to a reduction of about one year of education.

Residence in the Southern states is also shown to be associated with greater rates of English proficiency. The small positive influence of living in the South is similar to the effect recorded in Chiswick and Miller (2005a). As is apparent from the results presented in Table 1, columns (iii) and (iv), this finding only holds for immigrants from source countries other than Mexico.

Five further variables with behavioral interpretations that are constructed using birthplace information are included in the model – the miles of the origin country from the US, linguistic distance, minority language concentration and the colony and refugee variables. While the two language-related variables are constructed using information on the language other than English that is spoken in the immigrant's home, birthplace is also used to assign values for English-only speakers. The results show that English language skills improve with miles from the origin country, up to around 6,000 miles. The changes in English skills with miles from the origin country beyond this point to the maximum observed in the data (Indonesia, 8,985 miles), are relatively minor. Figure 5 portrays this relationship. To assist reading the figure a number of source countries are identified on the graph. As with Figure 4, the diagram has been calibrated so that the predicted relationship passes through the mean rate of English proficiency when the miles from the origin is at the sample mean (3,672 miles).

Figure 5
Predicted English Proficiency by Miles of Origin Country from the US, Adult Foreign Born Men by Origin, 2000 US Census



Source: Authors' calculations based on Table 1.

As argued when providing the conceptual background for the model of dominant language skills, the further the country of origin from the US the more intense the self-selection in migration and the less likely is return migration – both of which should be

associated with higher rates of English language proficiency. This is exactly the relationship depicted in Figure 5.

The measure of linguistic distance is included in the model as the reciprocal of the language scores presented in Chiswick and Miller (2005b). The language scores that form the basis for this measure range in value from 1.0 (Korean) to 3.0 (Swedish, Norwegian – see Table 1 of Chiswick and Miller (2005b)). As the reciprocal of the score is taken to form the linguistic distance variable, the variable in the model will range from 1 to $0.\dot{3}$, with a value of 1 indicating a home language quite far from English and $0.\dot{3}$ indicating a value close to English.[15]

The linguistic distance has a marked influence on English language proficiency. There is a potential difference of 0.66 in the linguistic distance measure, and this is associated with a difference of -0.861 in the probit index – an impact that is the equivalent of 8.1 years of education. In other words, the linguistic distance measure has a major impact on English proficiency. The predictions in Table 3 highlight this further.

Table 3
Predicted English Fluency by Linguistic Score, Adult Foreign Born Men by Origin.

			Predicted English Proficiency (%)	
Linguistic Score	Linguistic Distance	Illustrative Languages	Total Sample	Excluding Immigrants from Mexico
1.0	1.00	Korean, Japanese	44.01	57.65
1.5	0.67	Vietnamese, Arabic	61.20	74.49
2.0	0.50	Polish, Indonesia	69.22	81.36
2.5	0.40	Portuguese, Italian	73.65	84.87
3.0	0.33	Norwegian, Swedish	76.41	86.95

Source: Authors' calculations based on Table 1.
Note: All other variables evaluated at their means.

[15] The reciprocal of the original score was employed in the early research by Chiswick and Miller, as a means of assigning a score for English only speakers had not been developed and a zero value was used. In the current research, where the mean of the valid scores for the individual's birthplace is assigned where the individual speaks only English at home, there is less need to use the reciprocal functional form. While it is used here for consistency with past research, it is noted that entering the linguistic scores from Chiswick and Miller (2005b) in the probit index in linear form yields results similar to those listed in Table 1 and Table 6.

The minority language concentration variable in this first set of results is formed using State-level data (CONC-State). The results indicate that living in an area with a high representation of others who speak the same home language has a negative impact on an immigrant's English language proficiency. The minority language concentration variable ranges in value from 0 to 30 (percent), and so the variable, with a coefficient of -0.014, is associated with a change of up to 0.42 in the probit index. This is the equivalent of the impact on the probit index of four years of education. While this effect is far less than that associated with the linguistic distance measure, it is quite a pronounced effect, particularly considering that it is a neighborhood characteristic rather than an individual characteristic.

While the estimated impact of the minority language concentration variable is considerable, it appears to be less pronounced than that reported in the research based on the 1990 Census by Chiswick and Miller (2005a). Chiswick and Miller (2005a) estimate only a model that included birthplace fixed effects. Hence, to facilitate comparisons, a similar specification was estimated using the 2000 Census data. The estimated coefficient in this instance was -0.006. Comparison of this estimate with that presented in Chisiwck and Miller (2005a) needs to be sensitive to the changes between the 1990 and 2000 Censuses in both the rate of proficiency and the mean of the measure of minority language concentration. Hence, an elasticity is computed. The elasticity of English language proficiency with respect to the minority language concentration was 4.3 using the 1990 US Census (Chiswick and Miller (2005a)) and 3.0 using the 2000 US Census. Estimation of the model in Table 1 stratified by age and period of residence suggested the effect of a minority language concentration was greater (by about 15 percent) among those aged 35-64 years than among 25-34 year olds, and also among those resident in the US for 10 or more years (by about 45 percent) than for those who have resided in the US for less than 10 years. This suggests there has been a dilution of the negative impact of minority language concentration on English language proficiency among the more recent immigrants.

The final sets of variables in Table 1 record the influence on English language skills of being from a former British or US colony, or being a refugee. Coming from a former US/British colony is associated with higher rates of English proficiency, with the partial effect being a substantial 19 percentage points.[16] Being classified as a refugee is associated with poorer English skills, the partial effect being -8 percentage points. This is the equivalent of over two years of schooling.

Separate models were estimated in Table 1 for adult male immigrants from Mexico (column (iv)) and for all immigrants other than those from Mexico (column (iii)). For these disaggregated analyses, the estimating equations are modified as follows. First, for immigrants from Mexico the COLONY, REFUGEE and Linguistic Distance variables are not relevant.[17] Second, as noted above, the model for immigrants uses geographic distance variables defined with reference to the State of residence and Mexico City, Tijuana and Ciudaf Juarez. An alternative that was considered involved the use of dichotomous variables for US states that are near Mexico. Four groups are considered: California (benchmark); Texas; Arizona, New Mexico, Nevada; Rest of the US. Neither the geographic distance nor the State dichotomous variables have much explanatory power when the sample is confined to just immigrants from Mexico. The estimates presented are based on a geographic distance variable defined as the minimum of the distance between the capital of the immigrant's current state of residence and either Tijuana or Ciudaf Juarez, whichever is shorter. While the distance variables are insignificant, they are included for consistency with the other model specifications.

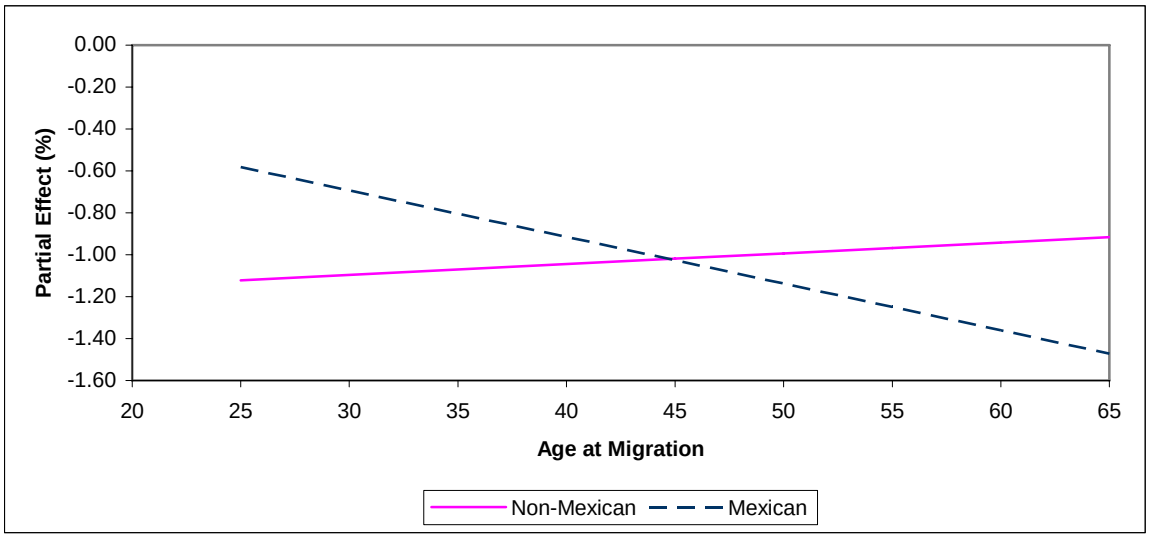
The estimates show that the effects discussed above are reasonably robust with respect to this disaggregation of the data between Mexico and all other countries. There are two exceptions though. First, the effects of age at migration on English proficiency follows different patterns, though in each instance the partial effect of age is negative across all

[16] Recall that immigrants from Canada, Australia, New Zealand and the English-speaking islands in the Caribbean have been deleted from the data set.

[17] While some immigrants from Mexico report speaking indigenous languages at home, there are no measures for the linguistic distance of Native American languages from English. Among the adult male immigrants from Mexico, 94.3 percent report speaking Spanish, 5.4 percent report only English and 0.3 percent all other languages.

relevant age groups. For immigrants other than those from Mexico, the linear age term is negative and the squared term positive. Both are statistically significant. The partial effect on English proficiency is around -1.04 percentage points at 40 years of age. For immigrants from Mexico, however, only the square of age is statistically significant. The partial effect of age on English proficiency for this group at 40 years of age is -0.92 percentage points, only marginally less than for the non-Mexican immigrants. With the estimated pattern of effects, however, it is apparent that migrating at an older age has a relatively larger negative effect on the English proficiency of immigrants from Mexico compared to the effect for other immigrants. This pattern is illustrated in Figure 6.

Figure 6
Partial Effect of Age at Migration on English Proficiency Among Adult Foreign Born Men, by Origin, 2000 US Census



Source: Authors' calculations based on Table 1.

The second difference between the results for immigrants from Mexico and all other immigrants is in the influence of children on the English skills of their fathers. While the effects of children on their father's English proficiency are always negative, these effects are much weaker in the disaggregated analysis, with the effect of the variable for living with children under 6 and 6-17 years being significant only for immigrants from Mexico. Among other immigrants the negative effect of children is significant only for the group of fathers with children only in the 6 to 17 age category. Note, however, that, taken as a

set the children variables are negative and statistically significant in both the Mexican and other analyses in Table 1.

Table 1, column (ii) lists results from the estimation of an ordered probit model with the full range of data on immigrants' language skills recorded in the LANG5 measure discussed above. The main feature of these results is that the estimated effects on the underlying ordered probit index are very much the same as for the index function in the binary probit model.

The effects of variables in the ordered probit model are difficult to assess, as the sign of a coefficient indicates unambiguously only the direction of the effect on the highest and lowest English proficiency groups. Accordingly, to illustrate the variation in membership of the five language categories distinguished in the LANG5 variable, predicted probabilities of being in a language category can be computed for the regressors included in the estimating equation. Sets of prediction are presented in Table 4 for education and duration of residence. These predictions have been computed using the mean values of all variables in the estimating equation other than for the specific variable that is the focus of the panels in the table.

The data in panel A of Table 4 show that increases in the level of education are associated with marked shifts away from the "not at all" and "not well" categories to the "very well" and "only English" categories. For education levels above the mean, there is also a shift away from the "well" category with more schooling. Comparison of the predicted distributions at 8 and 20 years of schooling shows that the change in proficiency, as measured by the binary scale adopted above, from 64 to 91 percent is also associated with quite pronounced shifts within the two groups "Not Proficient" and "Proficient".

Table 4
Predicted Distribution across Language Proficiency Categories by Level of Education and Duration of Residence, Adult Foreign Born Men, 2000.

A. Level of Education	Speaks Language Other than English at Home				Speaks only English	Total
	Not at All	Not Well	Well	Very Well		
8	7.51	28.54	34.61	26.67	2.68	100.00
10	5.41	24.57	34.66	31.47	3.90	100.00
12	3.80	20.62	33.80	36.24	5.54	100.00
14	2.60	16.87	32.12	40.73	7.68	100.00
16	1.74	13.44	29.72	44.69	10.40	100.00
18	1.14	10.44	26.79	47.87	13.76	100.00
20	0.72	7.91	23.52	50.06	17.79	100.00
B. Duration of Residence						
1	15.81	37.35	30.48	15.46	0.90	100.00
5	11.34	33.61	33.09	20.42	1.54	100.00
10	7.23	28.07	34.66	27.24	2.81	100.00
15	4.45	22.34	34.29	34.17	4.76	100.00
20	2.65	17.03	32.21	40.54	7.57	100.00
25	1.54	12.50	28.89	45.74	11.34	100.00
30	0.87	8.88	24.88	49.30	16.07	100.00

Source: Authors' calculations based on Table 1.
Note: Rows may not sum to 100.00 due to rounding. All other variables are evaluated at their means.

The data in panel A of Table 4 show that increases in the level of education are associated with marked shifts away from the "not at all" and "not well" categories to the "very well" and "only English" categories. For education levels above the mean, there is also a shift away from the "well" category with more schooling. Comparison of the predicted distributions at 8 and 20 years of schooling shows that the change in proficiency, as measured by the binary scale adopted above, from 64 to 91 percent is also associated with quite pronounced shifts within the two groups "Not Proficient" and "Proficient".

The predictions in Table 4, Panel B, for duration of residence, show a similar trend towards improvement in English skills with length of time in the US. Comparison of the

two extreme duration of residence levels represented in Table 4 indicate that the improvements in English skills with duration of residence are more pronounced than those associated with educational attainment.

In each case illustrated in Table 4, and in general, it is seen that the marked improvements in English proficiency associated with particular characteristics, such as educational attainment, duration of residence in the US and a younger age at migration, do not result in major changes in the proportion of immigrants becoming monolingual English speakers in the home. Rather, there is an extremely strong tendency for the mother tongue to be retained, so that most of the improvement in English proficiency is associated with increased representation in the category "Speaks a language other than English at home and speaks English Very Well". This phenomenon of mother tongue retention is analyzed in Section V.

Finally, each of the thresholds (μ's) in the ordered probit model is highly significant, indicating that the language categories coded in the Census are distinct, and as such analysis using an ordered polychotomous probit model may offer advantages over a study using the binary probit model. The similarity of the findings with the two approaches, and Kominski's (1989) caution that there is an absence of a clear differentiation between each of the four levels of English-speaking ability used in the US Census, suggests that the binary and polychotomous approaches to modeling English language proficiency offer similar findings.

Among the many strong influences on immigrants' English proficiency documented above, one that is of particular interest is associated with the minority language concentration variable. This variable captures an influence on English language skills of the "neighborhood" in which the immigrant lives. It was argued that living among others with whom the immigrant has a non-English language in common raises the cost of learning English and lowers the benefit from becoming proficient in English.

In the analysis reported in Table 1, the minority language concentration variable was constructed using state-level data. Two alternative geographic units are available for use in the construction of this variable: the Super Public Use Microdata Areas (CONC-Super) and the Metropolitan Statistical Areas (CONC-MSA). As discussed above, over 500 separate areas are identified with the Super-PUMA data, and over 100 with the MSA data, representing, respectively, a 10-fold increase and a doubling of the number of separate geographic entities compared to the State-level data. This more refined information may be important to use where the influences on English language proficiency associated with the neighborhood are more local than captured by the State-level aggregates.

Table 5 presents the coefficients for the alternative minority language concentration variables for the series of models presented in Table 1. There are two broad features of the results summarized in this Table. First, the impact of a minority language concentration on English language proficiency tends to be weaker when estimated with more disaggregated data than when state-level data are used, though the differences in the estimates do not affect in any way the overall interpretation that can be attached to this concentration or enclave influence. Second, the effects of a minority language concentration obtained with the more disaggregated data, reflecting the greater variability in the explanatory variable, appear to be estimated more precisely in the analyses disaggregated by birthplace region. With 't' values of 6 or more with the state-level data, however, obtaining even greater precision in estimation is not a major consideration. While it is not obvious which level of aggregation among the three is the "best" from an analytical point of view, and while it is not obvious which is the "best" from a measurement perspective (measurement errors), they all point to the negative association of ethnic/linguistic enclaves and immigrants' English language proficiency.[18]

[18] See Chiswick and Miller (2005c) for an "ethnic goods" model that hypothesizes this negative relationship.

Table 5
Estimates of Minority Language Concentration Variables, Language Models for Adult Foreign Born Men by Origin, 2000

Minority Language Concentration Variable: (Number of Areas)	Total Sample		Immigrants from All Countries except Mexico	Immigrants from Mexico
	Probit	Ordered Probit	Probit	Probit
State Data (51)	-0.0138 (18.75) [-0.0042]	-0.0163 (29.45)	-0.0135 (9.42) [-0.0027]	-0.0096 (6.24) [-0.0038]
MSA (106)	-0.0079 (18.57) [-0.0024]	-0.0116 (36.04)	-0.0086 (12.78) [-0.0017]	-0.0079 (10.57) [-0.0032]
Super PUMA (532)	-0.0060 (17.88) [-0.0018]	-0.0071 (27.79)	-0.0084 (14.43) [-0.0017]	-0.0041 (9.16) [-0.0016]

For notes to Table, see Table 1.

IV. ESTIMATES: FEMALE IMMIGRANTS

Table 6 presents estimates of the models of English language proficiency for females.[19] In cases where the estimated effects of a variable for females and males differ significantly, these are denoted by an asterisk against the estimated impact for females. These tests of statistical significance are based on equations estimated on a pooled sample of males and females with a full set of interaction terms for females. Tests were conducted of whether the data for males and females could be pooled and the effects on English proficiency represented by a single set of parameters common to both males and females. In each case the hypothesis of a common set of parameters was rejected.

In general, the direction of impact given by the estimated coefficients for female immigrants are the same as those established for males. For most variables the magnitudes of the estimated effects for females are of the same order as the estimated effects for males. Where significant differences arise, the differences (with the exception

[19] Females are generally shown to have an initial rate of learning advantage over males in first language acquisition. The limited evidence available suggests that this may carry over to second language acquisition. See, for example, Larsen-Freeman and Long (1991, pp. 204-205).

of the miles from the origin country variable) are small, though interesting. Five main differences arise.

Table 6
Probit Estimates of Language Models, Adult Foreign Born Women by Origin, 2000

Variables	Total Sample		Immigrants from All Countries except Mexico	Immigrants from Mexico
	Probit	Ordered Probit	Probit	Probit
Constant	-0.694 (7.24)	0.893 (13.26)	-0.096 (0.75)	-0.298 (1.74)
Education	0.114* (89.64) [0.037]	0.084 (91.97)	0.126* (72.51) [0.030]	0.086 (43.41) [0.034]
Age at Migration	-0.006 (1.30) [-0.011]	-0.011 (3.35)	-0.032 (5.52) [-0.011]	-0.009 (1.16) [-0.009]
Age at Migration Squared/100	-0.029 (5.53) [-0.011]	-0.017 (4.74)	-0.006 (0.90) [-0.011]	-0.016 (1.81) [-0.009]
Years Since Migration (YSM)	0.073 (43.19) [0.022]	0.054 (49.44)	0.086* (39.48) [0.020]	0.060* (21.60) [0.021]
YSM Squared/100	-0.041* (9.77) [0.022]	-0.020 (7.95)	-0.067 (12.24) [0.020]	-0.019* (2.87) [0.021]
Abroad 5 years ago	-0.238* (6.25) [-0.083]	-0.154* (5.50)	-0.333 (6.75) [-0.092]	-0.095* (1.59) [-0.037]
Married	-0.057* (4.88) [-0.018]	-0.057* (7.00)	-0.048* (3.10) [-0.011]	-0.054* (2.92) [-0.021]
With own children under 6 years only	-0.099* (5.28) [-0.033]	-0.101 (7.72)	-0.080 (3.32) [-0.020]	-0.068 (2.15) [-0.027]
With own children 6 to 17 years only	-0.116* (8.19) [-0.038]	-0.099* (10.09)	-0.110* (6.23) [-0.027]	-0.055 (2.27) [-0.022]
With own children under 6 years and 6 to 17 years	-0.209* (12.27) [-0.071]	-0.175* (14.32)	-0.166* (6.86) [-0.042]	-0.137* (5.26) [-0.054]
Non Metropolitan	0.004 (0.08) [0.001]	-0.016 (0.40)	0.552 (4.78) [0.097]	-0.092* (1.36) [-0.036]
South	0.123* (10.13) [0.039]	0.103 (12.24)	0.119 (7.36) [0.027]	0.095* (4.72) [0.038]
Miles (‘000) From Origin	0.315*	0.209*	0.297*	-0.696*

	(23.53) [0.023]	(22.69)	(18.45) [0.014]	(4.97) [-0.123]
Miles ('000) From Origin Squared	-0.031* (22.01) [0.023]	-0.025 (25.93)	-0.032* (19.65) [0.014]	0.192* (3.16) [-0.123]
Linguistic Distance	-1.204 (26.54) [-0.391]	-0.762* (24.22)	-1.282 (27.49) [-0.304]	(a)
Minority Language Concentration	-0.019* (23.57) [-0.006]	-0.022* (36.63)	-0.017* (12.01) [-0.004]	-0.021* (11.49) [-0.008]
Colony	0.830 (34.74) [0.215]	0.596 (43.73)	0.841 (33.86) [0.157]	(a)
Refugee	-0.224 (9.02) [-0.077]	-0.101 (5.48)	-0.203 (7.85) [-0.052]	(a)
μ_1	(a)	1.041 (217.58)	(a)	(a)
μ_2	(a)	1.848 (381.65)	(a)	(a)
μ_3	(a)	3.184 (457.46)	(a)	(a)
Chi-Squared	33990.68	37940.77	21084.55	5826.60
Prediction Success Rate	78.99	46.19	83.05	71.27
Sample Size	83832	83832	58000	25832

Source: 2000 US Census.
For notes to Table, see Table 1; * = estimate significantly different from that for males.

In general, the direction of impact given by the estimated coefficients for female immigrants are the same as those established for males. For most variables the magnitudes of the estimated effects for females are of the same order as the estimated effects for males. Where significant differences arise, the differences (with the exception of the miles from the origin country variable) are small, though interesting. Five main differences arise.

First, educational attainment is shown to have a slightly stronger (more positive) effect on the English skills of female immigrants than it has on the English skills of male immigrants. The differences in the partial effects, however, are all less than one percentage points. Educational attainment is included in the language model on the ground that it will capture efficiency factors. An implication of the current finding, therefore, is that these efficiency factors are more closely related to educational

attainment in the case of females than in the case of males. It may also capture labor supply effects as educational attainment is a far more important determinant of labor supply for women than for men.

The second variable where reasonably consistent differences emerge from the analyses for female and male immigrants is the "Lived Abroad 5 years Ago" variable. The sojourner behavior captured through this variable has a less negative impact on the English skills of female immigrants than it has on the English skills of their male counterparts. This difference may be a reflection of tied mobility, in which case the sojourner behavior would be more reflective of male intentions concerning length of stay in the US than it would be of female intentions.

Consistent with the discussion in Section II, children have a much more negative effect on the English skills of female immigrants than they have on the English skills of male immigrants. The partial effects for the samples pooled across birthplace regions are almost twice as large (in absolute value) for female immigrants as they are for male immigrants. Indeed, the negative effect of children on female English language proficiency is highly significant for all the children variables, even when separate analyses are done for Mexican and other immigrants.

The minority language concentration variable is also associated with a more negative impact on English language proficiency for female immigrants than it has for male immigrants. This finding could be associated with the lesser involvement of female immigrants in market work, an activity that is likely to offset some of the negative effects associated with living in a language enclave.

The final variable where a difference between males and females is apparent is for the geographic distance of the origin country from the US. The main difference here arises in the case of immigrants from Mexico. In the case of male immigrants from Mexico, the distance between the capital city of the immigrants' current state of residence and either Tijuana or Ciudal Juarez (whichever was the shorter) was not a significant determinant of

their English proficiency. Among female immigrants from Mexico, however, the geographic distance variable is statistically significant. Surprisingly, over the relevant range of distances represented in the data, the estimated coefficients indicate that immigrants living further away from the border with Mexico are less likely to be proficient in English than those who live close to the border. To put this another way, female immigrants living close to the border, for example in Texas, are more likely to be proficient in English than their counterparts who, say, live in Illinois.

When the geographic distance variable was replaced by a number of dichotomous variables for residence of a state that borders Mexico, similar findings were generated: the states that border Mexico are associated with much higher rates of English proficiency for female immigrants from Mexico than are states in the rest of the US. When similar variables for state of residence are entered into the estimating equation for male immigrants from Mexico, they are typically statistically insignificant, or where significant, are associated with very small impacts.

The reasons for this finding are unclear, and the most likely possibilities cannot be tested with the census data used in the current analysis. For example, it is possible that immigrants who live close to the border with Mexico originate from different regions of Mexico, possibly in border regions, compared to Mexican immigrants who live elsewhere in the US, who may originate from a broader catchment area. To the extent that they originate from border regions in Mexico, those resident in the states in the US near Mexico may have had greater exposure to English through more frequent trips to the US and US media prior to migration.[20]

V. ANALYSES OF MOTHER TONGUE RETENTION

The discussion of the results from the estimation of the ordered probit model showed that as years of residence in the US increase, immigrants' English language skills improve considerably. This improvement largely comes about through shifts to the better English

[20] The Census does not include information on the areas in Mexico in which the immigrants lived prior to migrating to the US.

ability categories ("well", "very well") from the poorer English ability categories ("Not at all", "Not well") among immigrants who speak a language other than English at home. While there is an important increase in the percentage of immigrants speaking only English at home, mother tongue retention is the dominant feature of the analysis of the five-category language proficiency variable (LANG5) above.

The aim of this section is to sharpen the focus on this issue of mother tongue retention by estimating language models where the dependent variable is a dichotomous variable (MT), set equal to one where the immigrant speaks a language other than English at home, and set equal to zero where the immigrant speaks only English at home. Around 90 percent of each of the samples of immigrants studied speak a language other than English at home. Even with such high rates of mother tongue retention, interesting patterns emerge from the analysis. Tables 7 and 8 list results from the model for males and females, respectively. The specification of the estimating equation that was used for the study of English skills in the previous section is employed in this analysis.

Table 7
Probit Estimates of Model of Mother Tongue Retention,
Adult Foreign Born Men by Origin, 2000

Variables	Total Sample	Immigrants from All Countries except Mexico	Immigrants from Mexico
	Probit	Probit	Probit
Constant	1.331 (11.78)	1.057 (7.51)	1.150 (5.15)
Education	-0.005 (3.49) [-0.001]	-0.015 (7.13) [-0.002]	0.014 (5.42) [0.002]
Age at Migration	-0.007 (1.30) [0.002]	-0.007 (1.15) [0.001]	0.014 (1.37) [0.000]
Age at Migration Squared/100	0.026 (4.21) [0.002]	0.034 (4.49) [0.001]	-0.018 (1.49) [0.000]
Years Since Migration (YSM)	-0.006 (3.64) [-0.004]	-0.015 (6.97) [-0.006]	0.009 (2.55) [0.000]
YSM Squared/100	-0.056	-0.052	-0.029

	(14.95) [-0.004]	(11.76) [-0.006]	(3.52) [0.000]
Abroad 5 years ago	0.035 (0.81) [0.005]	0.093 (1.67) [0.014]	-0.081 (1.21) [-0.009]
Married	0.141 (9.37) [0.021]	0.147 (7.89) [0.024]	0.096 (3.56) [0.010]
With own children under 6 years only	0.133 (5.92) [0.018]	0.184 (6.51) [0.027]	0.035 (0.90) [0.004]
With own children 6 to 17 years only	0.091 (5.23) [0.013]	0.102 (4.83) [0.016]	0.027 (0.83) [0.003]
With own children under 6 years and 6 to 17 years	0.149 (6.75) [0.020]	0.160 (5.46) [0.024]	0.092 (2.61) [0.009]
Non Metropolitan	-0.081 (1.38) [-0.012]	-0.376 (4.25) [-0.076]	0.033 (0.39) [0.003]
South	-0.041 (2.86) [-0.006]	-0.038 (2.13) [-0.006]	0.046 (1.68) [0.005]
Miles (‘000) From Origin	-0.226 (14.57) [0.009]	-0.095 (5.20) [0.021]	-0.138 (0.77) [0.004]
Miles (‘000) From Origin Squared	0.035 (21.06) [0.009]	0.025 (13.71) [0.021]	0.086 (1.08) [0.004]
Linguistic Distance	0.372 (6.11) [0.053]	0.359 (5.77) [0.057]	(a)
Minority Language Concentration CONC-STATE	0.020 (19.26) [0.003]	0.034 (17.57) [0.005]	0.000 (0.13) [0.001]
Colony	-0.677 (31.40) [-0.133]	-0.675 (30.63) [-0.136]	(a)
Refugee	-0.071 (1.77) [-0.011]	-0.091 (2.18) [-0.015]	(a)
Chi-Squared	6138.56	5755.66	104.86
Prediction Success Rate	90.77	88.65	94.71
Sample Size	85865	54001	31864

For notes to Table, see Table 1.

Many of the broad patterns from the study of mother tongue retention for males (Table 7) are the complement of those reported from the analysis of English proficiency. There are, however, a number of cases where the impact of variables on mother tongue retention

diverge from that which might have been expected given the results from the study of English proficiency.

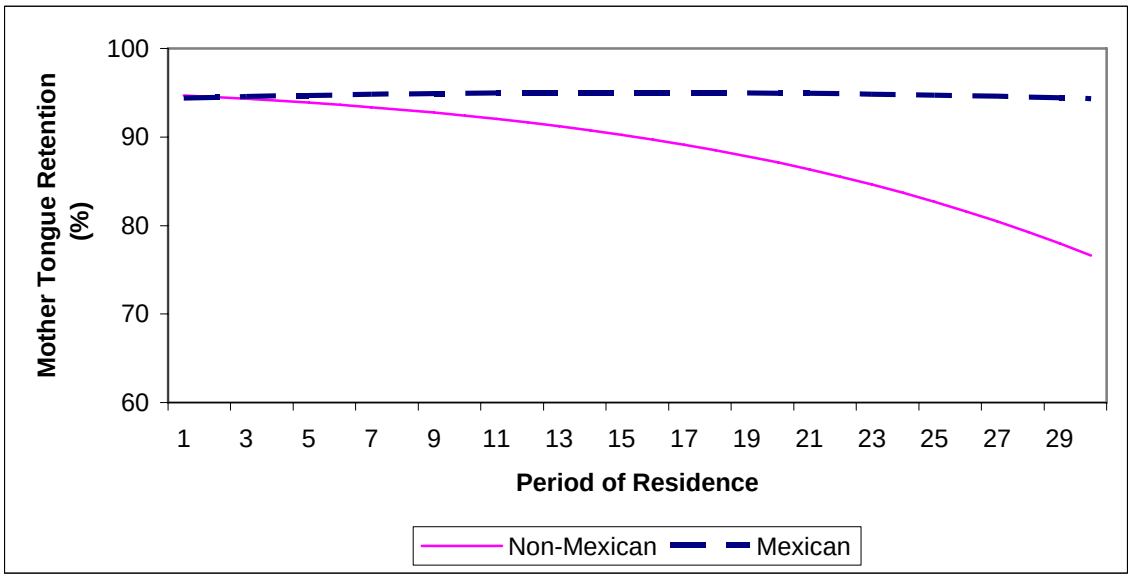
In the analyses pooled across birthplace regions, mother tongue retention is less likely among the better educated than it is among the less-well educated. However, in the separate analyses conducted for the broad birthplace regions, it is found that the impact of educational attainment differs markedly for immigrants from Mexico and for immigrants from countries other than Mexico. For male immigrants from countries other than Mexico, mother tongue retention is less likely among the better educated. In comparison, for male immigrants from Mexico, mother tongue retention is more likely among the better educated.

This difference could arise for a number of reasons. It could be associated with a greater likelihood of return migration for the better educated from Mexico compared to the less-well educated, relative to the pattern for immigrants from countries other than Mexico. A second possibility is that there is less inter-birthplace marriage among the better educated from Mexico than for the less-well educated. Finally, the finding could be a reflection of the different segments of the distribution of educational attainments that are relevant in each sample. Hence, the mean educational attainment of immigrants from countries other than Mexico is 13.3 years, and that of immigrants from Mexico is only 8.3 years. If this is the case, however, the educational attainment – mother tongue relationship for the full sample would have to be characterized by very intense non-linearities. Inclusion of an "education squared" term in the estimating equation for immigrants from countries other than Mexico did not reveal any evidence for a positive relationship between mother tongue retention and educational attainment among the less-well educated. However, in the analyses for Mexico, the mother tongue retention – educational attainment relationship was positive among the less-well educated (schooling levels up to 12.7 years for males, and up to 13.6 years for females) and negative among the better educated. This evidence offers support for the third proposition advanced above.

The impact of age at migration on mother tongue retention also differs between male immigrants from Mexico and those from other countries. For immigrants from countries other than Mexico, the older the age at migration, the more likely the immigrant is to retain his mother tongue. Age at migration, however, is not a significant determinant of mother tongue retention among immigrants from Mexico. Immigrants from Mexico are more likely to have lived abroad (presumably in Mexico) 5 years ago than are other immigrants. They presumably are characterized by much more to-and-fro migration between their country of origin and the US than are other immigrants. This could be associated with incentives to retain the origin country language that dominate the age at migration influences.

Similar to the patterns associated with education and age at migration, the duration of residence effects on mother tongue differ between immigrants from Mexico and immigrants from other countries. For immigrants from countries other than Mexico, the probability of retaining the mother tongue declines with duration of residence. The partial effect of duration of residence on the probability of speaking a language other than English at home is 0.4 percentage points when evaluated at 15 years of residence. Among immigrants from Mexico, however, there is virtually no change with duration of residence in the probability of speaking a language other than English at home, as illustrated in Figure 7.

Figure 7
Predicted Probability of Speaking a Language other than English at Home, Adult Foreign Born Men by Origin, 2000 US Census



Source: Authors' calculations based on Table 7.

The family structure variables have strong impacts on the probability of mother tongue retention. Mother tongue retention is greater among those who are married and who have children. However, the partial effects of these significant variables are double in magnitudes for Mexican immigrants compared to immigrants from other countries. This difference could arise from the importance of Spanish in the US (see Appendix Table A.1).[21] In this situation, where these general neighborhood influences appear so strong, the more immediate neighborhood effect associated with the family might be expected to have less influence.

In Table 7, column (i), for the pooled sample, the probability of mother tongue retention declines with the distance of the country of origin from the US up to around 3,500 miles, and increases thereafter. This result is reasonably consistent with the finding for the ordered probit model (Table 1), where the probability of speaking only English at home rises with geographic distance over the first 4,700 miles, and declines thereafter.

[21] In the US in 2000, fully 11 percent of individuals aged 18-64, and 60 percent of the foreign born of the same age group, speak Spanish at home. The next most frequently reported language is French, spoken at home by 0.6 percent of the adult population and by 3.4 percent of the foreign born.

The results for the linguistic distance variable indicates that immigrants whose mother tongue is closer to English (and hence who should find it easier to learn English) are most likely to shift from speaking their mother tongue to English. This finding is in the expected direction. Note that the sizes of the estimated partial effects are much smaller than the results in the study of English proficiency (Table 1). Given the motivation in the design of the linguistic distance variable, as a measure of how easy it is for foreign language speakers to learn English, the relative magnitudes of these estimated effects make sense.

The likelihood of immigrants speaking a language other than English at home rises with the extent to which the immigrant lives among others who speak the same (non-English) home language as the immigrant. However, this effect does not carry across to immigrants from Mexico.[22] This minority language concentration variable is intended to capture the ease with which conversation in a language other than English is possible. With the dominant position of Spanish in the US (see Appendix Table A.1) and the proximity to Mexico, and hence to Mexican media, this may not be as important a consideration as it is in the study of the incidence of mother tongue retention among non-Mexican immigrants.

Finally, it is seen that immigrants from a former colony of the US or UK, and those who are likely to have been refugees, are less likely to speak a language other than English at home. The latter may reflect a lower prospect of return migration.

Similar conclusions can be drawn from the study of mother tongue retention among adult female immigrants. These results, presented in Table 8, have four main features.

Table 8
Probit Estimates of Model of Mother Tongue Retention,
Adult Foreign Born Women by Origin, 2000

[22] The mean value of the minority language concentration (STATE) variable is 20.6 for immigrants from Mexico compared to a value of only 4.9 for other immigrants.

Variables	Total Sample	Immigrants from All Countries except Mexico	Immigrants from Mexico
	Probit	Probit	Probit
Constant	0.928 (8.25)	0.772 (5.74)	0.152 (0.64)
Education	0.008 (4.91) [0.001]	-0.003 (1.56) [-0.001]	0.023 (8.15) [0.003]
Age at Migration	0.000 (0.09) [0.002]	0.005 (0.74) [0.002]	0.002 (0.21) [-0.000]
Age at Migration Squared/100	0.016 (2.61) [0.002]	0.017 (2.37) [0.002]	-0.006 (0.52) [-0.000]
Years Since Migration (YSM)	-0.009 (5.25) [-0.004]	-0.016 (8.06) [-0.005]	0.009 (2.51) [0.000]
YSM Squared/100	-0.039* (10.63) [-0.004]	-0.037* (8.67) [-0.005]	-0.023 (2.71) [0.000]
Abroad 5 years ago	0.006 (0.13) [0.001]	0.065 (1.17) [0.010]	-0.154 (1.87) [-0.020]
Married	0.103* (7.61) [0.016]	0.147 (9.20) [0.025]	-0.320* (1.18) [-0.004]
With own children under 6 years only	0.108 (4.79) [0.016]	0.115 (4.36) [0.018]	0.044 (0.96) [0.005]
With own children 6 to 17 years only	0.109 (6.65) [0.016]	0.124 (6.48) [0.020]	0.023 (0.66) [0.003]
With own children under 6 years and 6 to 17 years	0.136 (6.36) [0.019]	0.159 (5.77) [0.024]	0.046 (1.22) [0.005]
Non Metropolitan	-0.035 (0.57) [-0.006]	-0.340 (4.13) [-0.069]	0.267 (2.58) [0.026]
South	-0.077* (5.50) [-0.012]	-0.076 (4.58) [-0.013]	-0.085* (2.93) [-0.010]
Miles (‘000) From Origin	-0.128* (8.18) [0.015]	-0.056 (3.06) [0.023]	1.172* (6.25) [0.056]
Miles (‘000) From Origin Squared	0.027* (16.25) [0.015]	0.021 (11.62) [0.023]	-0.358* (4.20) [0.056]
Linguistic Distance	0.024*	-0.002*	(a)

	(0.46) [0.004]	(0.03) [-0.000]	
Minority Language Concentration CONC-STATE	0.026* (24.45) [0.004]	0.035 (18.46) [0.006]	0.029* (11.99) [0.003]
Colony	-0.688 (32.40) [-0.142]	-0.696 (32.17) [-0.146]	(a)
Refugee	-0.009 (0.24) [-0.001]	-0.080 (2.05) [-0.014]	(a)
Chi-Squared	5192.54	5064.45	255.83
Prediction Success Rate	89.97	88.40	93.74
Sample Size	83832	58000	25832

First, the model estimated for females are, as a set, statistically different from the estimated for males.[23]

Second, the way educational attainment impacts on mother tongue retention differs for males and females. In particular, while mother tongue retention decreases with educational attainment among male immigrants from countries other than Mexico, educational attainment is not a significant determinant of mother tongue retention among female immigrants from countries other than Mexico. Like their male counterparts, the better educated female immigrants from Mexico are more likely to speak a language other than English at home than is the case for less-well educated female immigrants. However, also similar to the case for males, when the education variable is entered in the model in quadratic form, the results indicate that mother tongue retention initially increases with educational attainment, but after around 13.6 years of schooling, additional years of education are associated with decreases in the likelihood of a language other than English being spoken at home.

Third, neither the marital status variable nor the variable for the presence of children is a significant determinant of mother tongue retention among female immigrants from Mexico.

[23] The χ^2 test of whether the female shift variable and female interaction terms on all slope variables were jointly significantly different from zero was highly significant in each sample.

Fourth, among female immigrants from Mexico, the geographic distance variables are significant, whereas they were insignificant among male immigrants from Mexico. The estimated coefficients indicate that, over the range of distances represented in the sample for immigrants from Mexico, the likelihood of speaking a language other than English at home increases at a decreasing rate the further from Mexico the female immigrants live. While this is an unexpected finding, it sits comfortably alongside the finding reported earlier to the effect that the probability of being proficient in English decreases with distance from the border between US and Mexico.

Finally, the linguistic distance variable is not statistically significant in any of the models estimated for females. While this variable was statistically significant in the models for males, the estimated impacts were far less than those established in the model for English proficiency. The linguistic distance variable therefore appears to be more relevant to models capturing shifts to English than it is to models of origin language retention.

It is apparent from this discussion of the incidence of immigrants speaking a language other than English at home that the influences of both the personal characteristics and the behavioral variables included in the model vary between immigrants from Mexico and those from other countries. Given the extent to which Spanish is used in the US – as noted above, one in nine people between the ages of 18 and 64, and every second foreign born person in this age bracket, speaks Spanish at home – differences of this nature might be expected. However, the high degree of mother tongue retention among adult immigrants (of 90 percent or more) suggests that the appropriate research questions might be focused on the language use of the children of immigrants. Chiswick, Lee and Miller (2005a)(2005b) show how the family setting can be incorporated into the study of English speaking skills. Applying the same framework to the study of mother tongue retention is an appropriate direction for future research.

VI. SUMMARY AND CONCLUSION

One in nine people between the ages of 18 and 64 in the US, and every second foreign-born person in this age bracket, speaks Spanish at home. And whereas around 80 percent of adult immigrants in the US from non-English speaking countries other than Mexico are proficient in English, only about 50 percent of adult immigrants from Mexico are proficient.

A theoretical model, based in human capital theory, is presented in this paper to account for these empirical facts. Variations in English proficiency are shown to be linked to factors that shift the supply and demand for funds for investment in English language capital. The model has three key sets of factors: exposure, efficiency and economic incentives.

The exposure factors are separated into exposure to English before immigration (*e.g.*, living in a former US or British colony), and exposure after immigration. Exposure after immigration has both extensive and intensive margins. The extensive margin can be measured by the number of years since the immigrant arrived in the US, taking into account sojourner effects. The intensive margin is the intensity of exposure per unit of time in the US, and this will depend on the immigrant's neighborhood and family experiences. Special attention is given in the study to the most appropriate measure of neighborhood or enclave in the context of models of English proficiency.

Four measurable efficiency factors that can influence the development of language skills among immigrants are age at migration, educational attainment, refugee status, and linguistic distance. The conceptual basis for the measure of linguistic distance is the ease with which immigrants of particular linguistic backgrounds (mother tongues) can learn English. It is argued that it should be more difficult for a Chinese speaker to learn English than it is for a Spanish speaker to learn English because the differences between the languages are much greater in the former case than in the latter case. A measure of this "linguistic distance" is provided by Chiswick and Miller (2005b).

The economic factors included in the conceptual framework are the wage, consumption and social gains associated with English proficiency, including expected duration in the destination. Several proxies for these are considered, including educational attainment and the geographical distance between the country of origin and the US.

The empirical analyses are conducted using the 2000 US Census 1 percent Public Use Microdata file. The research is limited to adults (aged 25-64 years), though in contrast to most of the literature which focuses on males, the language practices of both male and female immigrants are considered. Binary probit analysis of a dichotomous proficiency index, and ordered probit analyses of a five-category language skills variable, are presented. As well, a binary measure of mother tongue retention is analysed.

The analyses show that immigration at a younger age, increases in the level of education and a longer duration in the US are associated with a higher probability of being proficient in English. English language proficiency tends to be lower the greater the age at migration. It is shown that among immigrants from countries other than Mexico, age 15 is a threshold in terms of the learning of English as a second language in the US. A similar threshold apparently does not exist among immigrants from Mexico, and this was attributed to the widespread use of Spanish as a home language.

English proficiency improves at a decreasing rate with duration of residence. Analyses performed for the 1980 and 1990 Censuses show a similar pattern, although it is noted that the effects of years since migration get slightly stronger across cohorts. That the pattern repeats itself suggests that it is reflecting a longitudinal phenomenon rather than merely a decline in the linguistic proficiency of more recent cohorts, or selective emigration of the least proficient immigrants in each arrival cohort. That the relationship gets stronger in the more recent data may be linked to the slightly lower measured proficiency in English of recent arrival cohorts: there is a well established pattern in the immigrant adjustment literature of faster rates of adjustment among those with the fewest skills at arrival.

Spending time abroad after immigration diminishes English proficiency. Thus, the intermittent nature of the stay in the US among those who came to the US to stay more than five years ago, but who lived abroad in 1995, and perhaps the expectation of a relatively short future stay in the US among them, appears to have a major negative influence on immigrants' English skills.

A greater geographic distance between the country of origin and the US was associated with greater proficiency in English, and a greater linguistic distance between the immigrant's mother tongue and English was associated with lesser proficiency in English. Both patterns are consistent with the theoretical model, and with results reported in previous research. The estimated effect is also very large: the estimated differences in English proficiency of the immigrants with mother tongues at the extremes of the measure of linguistic distance (*e.g.*, Korean, Japanese versus Norwegian, Swedish) is the equivalent of over eight years of schooling.

The proportion of individuals living in the same region as the immigrant that speak his or her mother tongue was also shown to have a major influence on the immigrant's English skills. Three measures of minority language concentration were used. The first was constructed using state-level data. As in previous research, 51 separate areas are used in the construction of this variable. The second was constructed using Metropolitan Statistical Areas (106 separate areas), and the third was based on the Super-PUMA (532 separate areas). The impact of a minority language concentration on English language proficiency tends to be weaker when estimated with more disaggregated data than when state-level data are used. It was not possible to assess which level among the three is the best from an analytical point of view, or even from a measurement perspective. However, the results with each of the three alternative measures point to a negative association of ethnic/linguistic enclaves and immigrants' English language proficiency.

The directions of impact of variables in the models of English skills for females were remarkably similar to those for males. The estimated coefficients for females as a set were, however, statistically different from those for males. Among the variables giving

rise to this statistically significant result are the minority language concentration variable and the variables for children. There is a more negative impact on English proficiency of living in a minority language concentration for female immigrants than for male immigrants. It was suggested that this difference could be associated with the lesser involvement of female immigrants in market work.

Children have a much more negative effect on the English skills of female immigrants than they have on the English skills of male immigrants. It was argued that this difference may arise because of the negative effect of children on mother's labor supply and children serving as translators for their mothers in consumption activities.

The analysis of mother tongue retention showed that many of the variables that affected immigrants' proficiency in English were significant determinants of their likelihood of retaining their mother tongue in the US. In many cases variables that are associated with greater proficiency (*e.g.*, years since migration, fewer children) are associated with a lower probability of speaking a language other than English at home. However, in some cases variables have impacts of the same sign on both English proficiency and mother tongue retention (*e.g.*, years of education among male immigrants from Mexico) or were significant in one set of analyses and insignificant in the other (*e.g.*, the linguistic distance variable for female immigrants). These differences indicate that additional information can be gained from the separate study of mother tongue retention.

The results from study of mother tongue retention, however, showed much more variability across birthplace groups and between males and females than was the case with the study of English proficiency. The differences between immigrants from Mexico and other countries are presumably associated with the dominant role of Spanish in the US. The differences between males and females could be associated with females having a more influential role than males in the transmission of the language and culture of the origin country within the family. Study of mother tongue retention using the family as the unit of observation may therefore be a priority in research. In this context, the very high rates of mother tongue retention among adult immigrants (of 90 percent or more)

suggests that the appropriate focus in this research would be on the home language use of the children of immigrants. This will provide information on the determinants of whether the use of languages other than English at home is more than a one-generational phenomenon.

There are two main lessons from this research. First, English proficiency varies across immigrants in predictable ways. Immigrants with higher educational attainment, who migrate at a younger age, who live in predominately English-speaking areas of the US, and who have a mother tongue close to English or who had been exposed to English prior to immigration are more likely to be proficient in English than other immigrants. These findings could be used in immigrant selection (*e.g.*, in a points system such as that used in Australia). The similarity of the estimated impacts across the separate groups analyzed above (males, females, immigrants from Mexico, immigrants from other countries) suggest that the application of the findings in this way would be neutral across these broad groups. The results also have implications for settlement policy, suggesting a focus on areas of ethnic language concentrations. Moreover, comparisons between men and women and with the 1980 and 1990 Census suggest that the results are robust across the two genders and across time.

Second, mother tongue retention also varies across immigrants in predictable ways. The overall rate of mother tongue retention among adult immigrants in the US is very high, at around 90 percent. Perhaps the more interesting issue that arises from this is the transmission of the mother tongue from adult immigrants to their children, and the labor market implications of bilingualism among the children of immigrants. These issues are the topics for future research.

REFERENCES

Becker, Gary S. and Chiswick, Barry R. (1966). "Education and the Distribution of Earnings", *American Economic Review*, Vol. 56, No. 2, pp.358-369.

Chiswick, B., (1978) 'The Effects of Americanization on the Earnings of Foreign Born Men', *Journal of Political Economy*, Vol. 86, pp.897-921.

Chiswick, B., (1979) 'The Economic Progress of Immigrants: Some Apparently Universal Patterns', in William Fellner, ed. *Contemporary Economic Prob*lems *1979*, American Enterprise Institute, Washington, pp.357-399.

Chiswick, Barry R. (1991). "Speaking, Reading and Earnings Among Low-Skilled Immigrants," *Journal of Labor Economics*, Vol. 9, No. 2, pp.149-70.

Chiswick, Barry R., Lee, Yew Liang and Miller, Paul W., (2005a). "Family Matters: The Role of the Family in Immigrants' Destination Language Acquisition", *Journal of Population Economics*, Vol. 18, No. 1, 2005, pp.1-17.

Chiswick, Barry R., Lee, Yew Liang and Miller, Paul W., (2005b). "Parents and Children Talk: English Language Proficiency Within Immigrant Families", *Review of Economics of the Household*, (forthcoming).

Chiswick, Barry R., Lee, Yew Liang and Miller, Paul W., (2005c) "Immigrants' Language Skills and Visa Category", *International Migration Review*, (forthcoming).

Chiswick, Barry R. and Miller, Paul W., (1992). "Language in the Immigrant Labour Market", pp.229-296 and pp.471-476 in Barry R. Chiswick (ed.), *Immigration, Language and Ethnicity: Canada and the United States*, American Enterprise Institute, Washington D.C..

Chiswick, Barry R. and Miller, Paul W., (1995). "The Endogeneity Between Language and Earnings: An International Analysis", *Journal of Labor Economics*, Vol. 13, No. 2, pp.246-288.

Chiswick, Barry R. and Miller, Paul W., (1996). "The Languages of the United States: Who Speaks What and What it Means", *READ Perspectives*, Vol. 3, No. 2, Fall 1996, pp.5-41.

Chiswick, Barry R. and Miller, Paul W., (1998). "English Language Fluency Among Immigrants in the United States", *Research in Labor Economics*, JAI Press, Vol. 17, pp.151-200.

Chiswick, Barry R. and Miller, Paul W., (2001). "A Model of Destination Language Acquisition: Application to Male Immigrants in Canada", *Demography*, Vol. 38, No.3, pp.391-409.

Chiswick, Barry R. and Miller, Paul W., (2002). "Immigrant Earnings: Language Skills, Linguistic Concentrations and the Business Cycle", *Journal of Population Economics*, Vol. 15, pp.31-57.

Chiswick, Barry R. and Miller, Paul W., (2005a). "Do Enclaves Matter in Immigrant Adjustment?", *City and Community*, Vol. 4, No. 1, pp.5-35.

Chiswick, Barry R. and Miller, Paul W., (2005b) "Linguistic Distance: A Quantitative Measure of the Distance Between English and Other Languages", *Journal of Multilingual and Multicultural Development*, Vol. 26, No. 1, pp.1-11.

Chiswick, Barry R. and Repetto, Gaston, (2001). "Immigrant Adjustment in Israel: Literacy and Fluency in Hebrew and Earnings", in Slobodan Djajic (ed.) *International Migration: Trends, Policy and Economic Impact*, New York: Routledge, pp. 204-228.

Fathman, Ann., (1975). "The Relationship Between Age and Second Language Productive Ability", *Language and Learning*, Vol. 25, No. 2, pp.245-253.

Fitzpatrick, Gary L. and Modlin, Marilyn J. (1986). *Direct-line Distances: United States Edition*, Metuchen, NJ: The Scarecrow Press Inc.

Hart-Gonzalez, Lucinda and Lindermann, Stephanie, (1993). "Expected Achievement in Speaking Proficiency, 1993" School of Language Studies, Foreign Service Institute, U.S. Department of State, Mimeographed.

Kominski, R., (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.

Larsen-Freeman, Diane and Long, Michael H. (1991). *An Introduction to Second Language Acquisition Research*, Longman, New York.

Long, Michael H. (1990). "Maturational Constraints on Language Development", *Studies in Second Language Acquisition*, Vol. 12, No. 3, pp.251-285.

McManus, Walter, Gould, William and Welch, Finis, (1983). 'Earnings of Hispanic Men: The Role of English Language Proficiency', *Journal of Labor Economics*, Vol. 1, No. 2, pp. 101-130.

Miller, Paul W. and Neo, Leanne, (2003). "Labour Market Flexibility and Immigrant Adjustment", *Economic Record*, Vol. 79, No. 246, pp.336-356.

Newport, Elissa L. (1990). "Maturational Constraints on Language Learning", *Cognitive Science*, Vol. 14, No. 1, pp.57-77.

Service, Elisabet and Clark, Fergus I. M., (1993). "Differences Between Young and Older Adults in Learning a Foreign Language," *Journal of Memory and Language*, Vol. 32, pp.608-623.

APPENDIX A

DEFINITIONS OF VARIABLES

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

Data Source: 2000 Census of Population, Public Use Microdata Sample, 1 percent sample of the foreign born, except where noted otherwise.

Definition of Population: Foreign-born men and women aged twenty-five to sixty-four, born in countries other than the main English-speaking countries (UK, Ireland, Canada, Australia, New Zealand and the English-speaking Caribbean), territories of the United States, at sea or born abroad of American parents. Only residents of the 50 States and the District of Columbia are considered.

Dependent Variables:

English Language Fluency (LANG2 and LANG5): LANG2 is set equal to one for individuals who speak only English at home, or if a language other than English is spoken in the home, who speak English either "very well" or "well." The variable is set to zero where a language other than English is spoken in the home and the respondent speaks English either "not well" or "not at all." LANG5 is a polychotomous dependent variable (five categories), defined to include all proficiency categories: (i) speaks English only at home; speaks a Language other than English at home and speaks English (ii) very well; (iii) well; (iv) not well; (v) not at all.

Mother Tongue (MT) This is a binary variable, set equal to one for individuals who speak a language other than English at home (assumed to be their mother tongue). It is set equal to zero for individuals who speak only English at home.

Explanatory Variables:

Minority Language Concentration (CONC): Each respondent is assigned a measure equal to the percentage of the population aged eighteen to sixty-four in the region in which he/she lives, who reports the same non-English language as the respondent. In the construction of this variable, only the twenty-five largest non-American Indian language groups nationwide and the top five Indian language groups are considered. Details are provided in Table A.1.

For the 18-64 year old group used in the compilation of Table A.1, these constitute 92 percent of all responses where a language other than English is used at home. Representation in the other language groups is so small numerically that the proportions are approximately zero, and this value is assigned. Those who reported speaking only English are assigned the mean value of the CONC measure for other-language speakers of their birthplace group.

Three separate regional classifications and hence measures of CONC are considered. CONC-State is based on the state (50 States and the District of Columbia) of residence—a measure that corresponds to that used by Chiswick and Miller (2005a). The second is based on the Super-PUMA data available on the 1 percent file (CONC-Super). 532 separate regions are distinguished in the empirical analysis. The third measure is an intermediate case, and is constructed using the information on Metropolitan Statistical areas (CONC-MSA).[24] 106 separate regions are used in forming this measure.

Location: The three 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. In the analyses for Mexico, four dichotomous variables for US states that are near Mexico are used: California (benchmark); Texas; Arizona, New Mexico, Nevada; Rest of the US.

Colony (COLONY): Countries that are current or former colonies of English-speaking countries are coded one. All other countries are coded zero. Dependencies of the UK, US, Australia, New Zealand and South Africa are coded as colonies under this definition.

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

Lived Abroad Five Years Ago (ABROAD5): This is set equal to one if the individual had resided in the US for more than 5 years and lived abroad in 1995, otherwise it is set equal to zero. Note that ABROAD5 is zero for immigrants in the US for five or fewer years.

Marital Status (MARRIED): This is a binary variable that distinguishes individuals who are married, spouse present (equal to 1) from all other marital states.

Years of Education (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).

[24] A metropolitan area is one of a large population nucleus, together with adjacent communities that have a high degree of economic and social integration with that nucleus. Where a metropolitan area has 1 million people or more, two or more primary metropolitan statistical areas (PMSAs) may be defined within it. Information on the metropolitan statistical areas (MSA) and primary metropolitan statistical areas contained in the variable MSA-PMSA1 is used in the construction of the CONC-MSA variable.

Refugee (REFUGEE): This variable is constructed to identify the major sources of post-WWII refugees to the U.S. It is defined only for immigrants who migrated at age 25 and older. Individuals who migrated from Cambodia, Laos or Vietnam in 1975 or later, Iran in 1980 or later, Cuba in 1960 or later, the Former Soviet Republic in 1950 or later, from China between 1950 and 1990, Somalia or the former Yugoslavia between 1990 and 2000 are assigned a value of one for this variable. All other immigrants are assigned a value of zero.

Linguistic Distance (DISTANCE): This is a measure of the difficulty of learning a foreign language for English-speaking Americans. It is based on a set of language scores (LS) measuring achievements in speaking proficiency in foreign languages by English-speaking Americans at the U.S. Department of State, School of Language Studies, reported by Hart-Gonzalez and Lindermann (1993). It is described in detail in Chiswick and Miller (2005b).

In the construction of this variable, foreign-born persons who speak only English at home and hence do not report speaking a non-English language are assigned the mean value of the linguistic score measure for individuals reporting a foreign language from their birthplace group.

The variable in the regression equations is linguistic distance, which is one divided by the linguistic score, DISTANCE = 1/LS.

Direct-Line Distances (MILES): The miles between the major city in the immigrant's country of origin and the nearest large port of entry in the United States (New York, Miami, Los Angeles) are constructed from data in Fitzpatrick and Modlin's (1986) *Direct Line Distances, United States Edition*.

Means and standard deviations for these variables are reported in Table A.2 (for men) and Table A.3 (for women).

Table A.1
Frequency of Language Use Among Adults in the United States, 2000[(a)]

Language	% of All Languages		% of All Languages other than English	
	Frequency	Cumulative Frequency	Frequency	Cumulative Frequency
Spanish	11.228	11.228	59.743	59.743
French	0.632	11.860	3.364	63.107
Chinese, Min	0.623	12.483	3.316	66.423
Tagalog/Filipino	0.557	13.040	2.962	69.385
German , Austrian, Swiss	0.505	13.545	2.689	72.074
Vietnamese	0.433	13.978	2.302	74.376
Korean	0.383	14.361	2.038	76.414
Italian	0.329	14.690	1.751	78.165
Russian	0.274	14.964	1.459	79.624
Arabic	0.258	15.222	1.371	80.995
Portuguese	0.252	15.474	1.343	82.338
Polish	0.226	15.700	1.202	83.540
Japanese, Ainu	0.185	15.885	0.983	84.523
French Creole, Haitian Creole	0.175	16.060	0.932	85.455
Hindi	0.150	16.210	0.800	86.255
Persian, Dari, Farsi, Pushto	0.136	16.346	0.725	86.980
Greek	0.135	16.481	0.719	87.699
Urdu	0.109	16.590	0.580	88.279
Cantonese, Toishan	0.106	16.696	0.562	88.841
Gujarathi	0.099	16.795	0.529	89.370
Kru, Ibo, Yoruba, Akan, Ashanti, Ewe, Fanti, Ga, Igbo, Nigerian, Twi	0.086	16.881	0.457	89.827
Hebrew	0.083	16.964	0.442	90.269
Mandarin	0.077	17.041	0.408	90.677
Mon-Khmer, Cambodian, Khmer	0.069	17.110	0.365	91.042
Armenian	0.069	17.179	0.365	91.407
Navajo*	0.073	17.252	0.386	91.793
Dakota, Assiniboine, Lakota, Oglala, Sioux*	0.008	17.260	0.042	91.835
Cherokee*	0.008	17.268	0.041	91.876
Yupik*	0.006	17.274	0.033	91.909
Apache*	0.006	17.280	0.032	91.941
Other languages specified	1.514	18.794	8.056	99.997
Other languages not specifically listed	0.001	18.795	0.004	100.000
Not reported	0.000	18.795	0.002	100.000
English Only	81.206	100.000	-	-

Note: [(a)] All persons, regardless of nativity, aged 18 to 64 as reported in the 2000 Census; Refers to language spoken in the home if English is not the only language; * = American Indian Language
Source: 2000 US Census of Population, Public use Microdata Sample, 1 Percent Sample.

Table A.2
Means and Standard Deviations of Variables in Language Models, Adult Foreign Born Men by Origin, 2000

Variables	Total Sample	Immigrants from All Countries except Mexico	Immigrants from Mexico
Proficient in English (LANG2)	0.695 (0.46)	0.800 (0.40)	0.507 (0.50)
Education	11.525 (4.96)	13.323 (4.20)	8.279 (4.55)
Age at Migration	40.053 (10.30)	41.506 (10.47)	37.429 (9.44)
Age at Migration Squared	1710.296 (882.51)	1832.273 (911.21)	1490.107 (781.40)
Years Since Migration (YSM)	16.135 (11.04)	16.413 (11.51)	15.634 (10.13)
YSM Squared	382.307 (472.27)	401.889 (503.57)	346.956 (407.38)
Abroad 5 years ago	0.023 (0.15)	0.020 (0.14)	0.028 (0.16)
Married	0.622 (0.48)	0.644 (0.48)	0.583 (0.49)
With own children under 6 years only	0.129 (0.33)	0.119 (0.32)	0.147 (0.35)
With own children 6 to 17 years only	0.259 (0.44)	0.251 (0.43)	0.274 (0.45)
With own children under 6 years and 6 to 17 years	0.152 (0.36)	0.107 (0.31)	0.233 (0.42)
Non Metropolitan	0.010 (0.10)	0.005 (0.07)	0.020 (0.14)
South	0.282 (0.45)	0.273 (0.45)	0.299 (0.46)
Miles (‘000) From Origin	3.672 (2.64)	4.906 (2.56)	0.701 (0.39)
Miles (‘000) From Origin Squared	20.443 (23.11)	30.614 (23.24)	0.644 (0.81)
Linguistic Distance	0.508 (0.13)	0.543 (0.15)	(a)
Minority Language Concentration CONC-STATE	10.447 (11.10)	4.798 (7.92)	20.645 (8.40)
Minority Language Concentration CONC-MSA	12.880 (16.13)	7.369 (14.38)	22.829 (14.24)
Minority Language Concentration CONC-SUPER	15.746 (19.64)	8.799 (16.13)	28.287 (19.19)
Colony	0.150 (0.36)	0.233 (0.42)	(a)
Refugee	0.053 (0.22)	0.082 (0.27)	(a)

Sample Size	85865	54001	31864

Source: US Census of Population, 2000, Public Use Microdata Sample, 1 Percent Sample.
Note: Figures in parentheses are standard deviations; (a) Variable not relevant.

Table A.3
Means and Standard Deviations of Variables in Language Models, Adult Foreign Born Women by Origin, 2000

Variables	Total Sample	Immigrants from All Countries except Mexico	Immigrants from Mexico
Proficient in English (LANG2)	0.666 (0.47)	0.762 (0.43)	0.440 (0.50)
Education	11.483 (4.73)	12.815 (4.12)	8.344 (4.60)
Age at Migration	41.074 (10.62)	42.157 (10.69)	38.523 (9.99)
Age at Migration Squared	1799.821 (922.16)	1891.493 (938.83)	1583.786 (843.12)
Years Since Migration (YSM)	16.573 (11.46)	16.762 (11.76)	16.127 (10.70)
YSM Squared	405.906 (501.44)	419.201 (519.07)	374.577 (455.67)
Abroad 5 years ago	0.018 (0.13)	0.017 (0.13)	0.020 (0.14)
Married	0.670 (0.47)	0.672 (0.47)	0.667 (0.47)
With own children under 6 years only	0.120 (0.33)	0.116 (0.32)	0.130 (0.34)
With own children 6 to 17 years only	0.296 (0.46)	0.279 (0.45)	0.338 (0.47)
With own children under 6 years and 6 to 17 years	0.155 (0.36)	0.107 (0.31)	0.269 (0.44)
Non Metropolitan	0.009 (0.09)	0.006 (0.07)	0.017 (0.13)
South	0.277 (0.45)	0.276 (0.45)	0.280 (0.45)
Miles (‘000) From Origin	3.883 (2.65)	4.919 (2.53)	0.650 (0.35)
Miles (‘000) From Origin Squared	22.086 (23.29)	30.575 (23.04)	0.546 (0.73)
Linguistic Distance	0.520 (0.15)	0.552 (0.16)	(a)
Minority Language Concentration CONC-STATE	9.860 (11.04)	4.803 (7.91)	21.778 (7.70)
Minority Language Concentration CONC-MSA	12.670 (16.76)	7.433 (14.48)	25.012 (15.21)
Minority Language	15.526	8.763	31.466

Concentration CONC-SUPER	(20.28)	(16.10)	(20.18)
Colony	0.160	0.228	(a)
	(0.37)	(0.42)	
Refugee	0.056	0.080	(a)
	(0.23)	(0.27)	
Sample Size	83832	58000	25832

Source: US Census of Population, 2000, Public Use Microdata Sample, 1 Percent Sample.
Note: Figures in parentheses are standard deviations; (a) Variable not relevant.