DISCUSSION PAPER SERIES

IZA

IZA DP No. 2575

The Critical Period Hypothesis for Language Learning: What the 2000 US Census Says

Barry R. Chiswick
Paul W. Miller

January 2007

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

The Critical Period Hypothesis for Language Learning: What the 2000 US Census Says

Barry R. Chiswick
University of Illinois at Chicago and IZA

Paul W. Miller
University of Western Australia and IZA

Discussion Paper No. 2575
January 2007

IZA

P.O. Box 7240
53072 Bonn
Germany

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

IZA Discussion Paper No. 2575
January 2007

ABSTRACT

The Critical Period Hypothesis for Language Learning: What the 2000 US Census Says*

A critical period for language learning is often defined as a sharp decline in learning outcomes with age. This study examines the relevance of the critical period to English proficiency among immigrants in the US. It uses microdata from the 2000 US Census, a model of language acquisition from the economics and sociology literatures, and a flexible specification of an estimating equation based on 64 age-at-migration dichotomous variables. It shows that self-reported English language speaking proficiency among immigrants declines more-or-less monotonically with age at migration, and this relationship is not characterized by any sharp decline or discontinuity that might be considered consistent with a "critical" period. The findings are robust across the various immigrant samples, and between the genders.

JEL Classification: F22, J15

Keywords: immigrants, second language learning, critical period hypothesis

Corresponding author:

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

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

January 2007

THE CRITICAL PERIOD HYPOTHESIS FOR LANGUAGE LEARNING: WHAT THE 2000 US CENSUS SAYS

I. INTRODUCTION

It is generally agreed that learning a language is easier for younger than older people, where the measure of success is ultimate achievement (see, for example, Larsen-Freeman and Long, 1991, Scovel, 2000, Hyltenstam and Abrahamsson, 2003).[1] Many studies of age-related language acquisition have hypothesized a critical learning period. At ages below the critical period language learning can lead to native-like proficiency, or language acquisition can occur simply through exposure rather than through tutoring, or language skill acquisition is simply easier. At ages above the critical period learning a second language is much more difficult. A critical period should, according to Hakuta, Bialystok and Wiley (2003, p.31), be associated with "a significant change in learning outcome, not merely a monotonic decline with age". The empirical research on age-related differences has examined both first language and second language acquisition: A critical period in first language (mother tongue) acquisition does not necessarily imply a similar period in second language acquisition (Bialystok (1997, p.118)).

The literature has provided a number of theoretical models for the age-related differences, using explanations that are largely biologically-based (the brain's loss of plasticity) and cognitive-based (where problem solving and other learning approaches decline in effectiveness gradually with age).[2] The distinguishing feature of the cognitive-based explanations is that language proficiency declines reasonably smoothly with the age at which learning commences, whereas the biological-based accounts have a markedly different relationship between the age at which learning or exposure commences and the long-run attainment. That is, there is a discontinuity in the negative relation between age and language learning.

[1] Older learners may be at an advantage in some areas (*e.g.*, syntax), but this advantage is typically short-lived.

[2] See Hakuta, Bialystok and Wiley (2003) for a review of age-related changes in cognitive processes.

There have been many attempts to demark the critical period. Arriving at consensus findings from this research is difficult, however, as "…there is great variation among researchers on which age spans they use to divide up their subjects, and there may be multiple critical periods at varying age levels for different linguistic modalities…" Scovel (2000, p.215). Thus, a reasonably wide range of estimates have been advanced in the literature: for example, up to around 9 years by Penfield and Roberts (1959) and ages 2 to puberty by Lenneberg (1967). Krashen (1973) argues for a terminal age earlier than puberty. While most authors appear to subscribe to an upper critical period age between 5 to 15, some reject the notion of a critical period—see, in particular, Bialystok and Hakuta (1999), Hakuta, Bialystok and Wiley (2003) and Wiley, Bialystok, and Hakuta (2004).

Critical periods have been assessed using three types of studies: individual case studies, experimental laboratory studies, and analyses of large survey/census data sets. The focus in the case studies and experimental laboratory work has largely been on whether second language learners achieve native-like outcomes, and detailed test instruments have been employed. Both individual case studies and experimental laboratory studies have been based on small, select, samples.[3]

In contrast to the relatively small sample sizes involved in the case and laboratory studies, Bialystok and Hakuta (1999) provide a study of cross-tabulations from the 1990 US Census covering tens of thousands of observations.[4] They focus on Spanish and Chinese speaking immigrants in New York, with ages of arrival (or ages of onset) between zero and 70+, who have 10 years or more residence in the US (or equivalently, exposure to English). The analyses were conducted disaggregated by level of education, with educational attainment being shown to be positively related to English proficiency.

[3] See Bialystok (1997) for a critique of the evidence that supports the critical period hypothesis.

[4] While the US Census data contains only self-reported English proficiency for those who use a language other than English at home, and the cross-tabulations used by Bialystok and Hakuta (1999) report the age and age-at-arrival information only in intervals, these data have the advantage of a very large sample. Specifically, they analyzed data on 38,787 speakers of Spanish and 24,903 speakers of Chinese. In their subsequent study (Hakuta, Bialystok and Wiley (2003)), the sample sizes for these two groups were 2,016,317 and 324,444, respectively.

They show that English proficiency declines more or less continuously with age at arrival. This evidence was interpreted as rejecting the critical period hypothesis.

Hakuta, Bialystok and Wiley (2003) extend the analyses of the 1990 US Census data by Bialystok and Hakuta (1999), largely though covering additional States, and by a more rigorous statistical testing of the model.[5] A modified set of analyses is presented in Wiley, Bialystok and Hakuta (2004), in response to a critique offered by Stevens (2004). Specifically, Hakuta, Bialystok and Wiley (2003) test statistically for changes in the mean English proficiency, and also for changes in the partial effect of age, at two possible critical points, immigration at age 15 and at age 20. A non-parametric (local regression) approach was also employed. In each instance the authors report evidence of gradual declines in English proficiency with age at arrival in the US: no evidence was found of a discontinuity "that is the essential hallmark of a critical period" (Hakuta, Bialystok and Wiley (2003, p.37)). Hakuta, Bialystok and Wiley (2003) thus reject the critical period hypothesis. Wiley, Bialystok and Hakuta (2004) use the one percent Public Use Microdata Sample from the 1990 Census to provide analyses by region of birth as well as by language spoken at home. They report that their earlier conclusions on the lack of support for the critical period hypothesis carry over to these analyses.

The various studies by Bialystok, Hakuta and Wiley, being based on US Census data, have many similarities with the research of sociologists and economists on the determinants of English language proficiency (see, for example, Stevens (1992), Espenshade and Fu (1997) and Chiswick and Miller (1995)(2007a)). The distinguishing feature of the multivariate analyses conducted in these studies is the rich array of standardizing variables included in the statistical model, a feature intended to control statistically for a wide range of factors that influence language attainment. Hyltenstam and Abrahamsson (2003, p.559), for example, argue that "Among other factors, frequency and quality of input as well as identity issues seem to play an important role and interact with maturational constraints for the outcome even at a low age".

---

[5] They focus on California, Illinois, Texas and New York in the study of Spanish speakers, and California, Illinois, Texas, New York, Florida, Maryland, Massachusetts, New Jersey, Pennsylvania, Virginia and Washington in the study of Chinese speakers.

As some of the additional variables considered in the research by Chiswick and Miller (1995)(2007a), Espenshade and Fu (1997) and Stevens (1992) may be related to age at immigration, their absence in the study by Hakuta, Bialystok and Wiley (2003) may either accentuate or attenuate the age at immigration effects on ultimate English proficiency. Thus, one of the contributions of the current paper is to ascertain the relevance of the critical period hypothesis within the more encompassing framework these authors propose. In addition, being based on data from the 2000 US Public Use Microdata Sample, which contains information on both age and time in the US in individual years rather than as intervals, it avoids one of the problems (the use of broad categories for key variables) that Stevens (2004) has raised as a criticism of Hakuta, Bialystok and Wiley (2003), and which Wiley, Bialystok and Hakuta (2004) were not able to fully overcome. Finally, an alternative approach to examining the critical period hypothesis is outlined. This approach offers a flexible, and intuitive, way of identifying any critical period in the age of arrival-English proficiency relationship.

The structure of the paper is as follows. Section II outlines the approach taken in the recent work by Chiswick and Miller (1995, 2007a). Developed from an economics perspective, the model they propose has parallels with work in sociology (see, for example, Espenshade and Fu (1997) and Stevens (1992)), and is more encompassing than the model applied by Hakuta, Bialystok and Wiley (2003). Section III provides information on the data set used in the empirical work, the one percent Public Use Microdata Sample from the 2000 US Census. Section IV contains a brief discussion of the general findings on the determinants of self-reported proficiency in spoken English among immigrants from non-English speaking countries in the US. Most attention is devoted, however, to the links between self-reported proficiency in spoken English and age at migration. A feature of this analysis, however, is that by necessity, given the language questions in the Census, it focuses on only one dimension of language proficiency, self-reported proficiency in spoken English. Tests of robustness are offered in Section V, by way of separate analyses for immigrants according to whether their mother tongue is close to, or distant from, English. This partition of the sample seems particularly apt when studying English language proficiency. Section VI offers concluding comments.

II. AN ENCOMPASSING ECONOMIC MODEL OF ENGLISH PROFICIENCY

Chiswick and Miller (1995)(2007a) develop a conceptual framework that is used to account for variations in the English language proficiency of immigrants from non-English-speaking countries.[6] It is organized around three broad sets of influences: exposure, efficiency, and economic incentives.

The exposure variables take into account the environmental factors discussed by Hyltenstam and Abrahamsson (2003). Exposure to English 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. However, with advances in telecommunications and the world-wide spread of American and British media and movies, and more recently the internet, and the increasing practice of teaching English in the schools, 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 is used to capture some of these influences.

Post-immigration experience has been captured using information on the number of years since the immigrant came to the US to stay (an extensive margin) and by the immigrant's post-arrival neighborhood and family experiences (intensive margins). A quadratic specification for duration of residence 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, as is implied by human capital theory. This provides for a statistical control similar to that used by Bialystok and Hakuta (1999), who limit their analyses to those with 10 or more years of residence in the US, in order to focus on those whose language attainment process had stabilized.

The years since migration 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

[6] This model has been applied successfully to study the destination language proficiency of immigrants in the US, Canada, Australia and Israel. See Chiswick and Miller (2007b).

immigrants who arrived before 1995. A dichotomous variable is included in the model, which is unity where the immigrant came to the US to stay more than five years ago but lived abroad five years ago, and is zero otherwise.

The focus of the neighborhood variables in this research has been on the role of ethnic enclaves. Exposure to English is reduced by living and working in an area in which many others use his or her origin language. Reduced exposure results in reduced language learning. Measures of the linguistic concentration of individuals with whom the immigrant shares a mother tongue have been developed by Chiswick and Miller (1995) (2007a), and are incorporated into the estimating equation in the form of a measure of minority language concentration.[7]

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. It is not possible in the 2000 Census data to distinguish between pre- and post-migration marriages, but a marital status variable is included in the model.[8] In addition, the presence of children in the household could have a range of effects on immigrants' dominant language proficiency, including arguments based on children being teachers of their parents (positive effect on language learning), children being translators for their parents (negative effect), through children affecting the labor supply of their parents (negative effect on mothers), and parents seeking to transmit the culture of their country of origin, and in doing so they may encourage the learning of their origin language and reduce exposure to the destination language (negative effect), among their children. The Census data contain information on the presence and age structure of children, and this information is used in all models estimated.

---

[7] The minority language concentration variables may not capture the intended influences outside metropolitan areas (see Chiswick and Miller (1995)(2007a) for discussion). Regional and size of place dichotomous variables are also included in the estimating equation.

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

Efficiency refers to the ability to convert exposure into language learning. 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 age at migration effects have been discussed above.

The education effects are argued to arise from the better educated having technically superior language production functions, either because of greater innate learning ability or other unmeasured variables that enhance both formal education and language proficiency. 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. Bialystok (1997, p.130) points out that "…aspects of a second language that are structurally different from that of the first language are more difficult for learners to master".[10] 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 for estimation purposes by Chiswick and Miller (2005) and Ginsburgh, Ortuño-Ortín and Weber (2005). The

---

[9] That education has a significant effect on Hebrew language proficiency among immigrants in Israel suggests that pre-immigration exposure to the destination language in school is not the main effect of education.

[10] Bialystok (1997, p.130) argues that "…first language acquisition is fundamental and guides and influences second language acquisition. The specific structures that transfer across languages, the nature of that transfer and the implications of transfer for mastering a second language are issues that need to be pursued in further studies".

Chiswick and Miller (2005) 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 Ginsburgh *et al*. (2005) measure is based on data collected by Dyen on Indo-European languages and is described in Dyen, Kruskal and Black (1992). As the Chiswick and Miller measure covers a greater range of mother tongues and not merely Indo-European languages, and is based on learning behaviour, it is employed in the current study.

Admission criteria may be relevant for understanding immigrant linguistic adjustment. 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 or social adjustment. Also, refugees often have less time to prepare for the move. A reasonable refugee variable can be constructed using information on country of birth, period of immigration, and age at migration (see Cortes (2004)). The latter criterion permits refugee status to influence dominant language outcomes only where the person entered the US as an adult.[11]

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. Two sets of factors will be important: the expected improvements in economic outcomes (*e.g.* increments in wages) from becoming proficient in the dominant language, and the length of time these benefits are likely to be received. As there are strong links between educational attainment and the economic (wage) returns from becoming proficient in the

[11] 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.

dominant language for immigrants (see, for example, Chiswick and Miller (2003)), the individual's level of education may serve as a proxy for the expected economic wage returns from the investment in dominant language skills.

The length of time the benefits associated with the improvements in language skills are received will be affected by the probability of return migration. 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 a greater incentive to become proficient in English. This distance effect is captured through a variable for the number of thousands of miles from the major city in the origin country to New York, Miami or Los Angeles, whichever is the shorter.[12] (See Espenshade and Fu (1997) and Chiswick and Miller (1995)(2007a) for further details.)

Hence, the empirical counterpart of the conceptual framework developed by Chiswick and Miller (1995)(2007a) 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)

The variables and mnemonics are defined in Table 1.

Table 1
Definitions of Variables Used in the Statistical Analysis of English Proficiency Among Immigrants from Non-English Speaking Countries, 2000 US Census

Variable Name/ Mnemonics	Description
Educational Attainment	Years of education.
Age at Migration	Age minus years of residence in the US.
YSM (Years Since Migration)	The number of years since the immigrant first arrived in the US to

[12] When the analysis is limited to immigrants from Mexico, the geographic distance variable is computed as the minimum of the direct line distance between the capital of their current state of residence and either Tijuana or Ciudad Juarez.

	stay.
YSMSQ	The square of YSM.
ABROAD*	Lived abroad five years ago for those who immigrated more than 5 years ago.
MARRIED*	The individual is married, with spouse present.
Children*	Has a child aged under 6; Has a child aged 6-17; Has children under 6 and 6-17 years of age.
NON-MET*	Lives outside the metropolitan areas.
SOUTH*	Lives in Alabama, Arkansas, Delaware, District of Columbia, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, Missouri, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia or West Virginia.
MILES	Direct line distance between the major city in the immigrant's country of origin and either New York, Miami, or Los Angeles (whichever is the shortest), in thousands of miles.
MILESQ	The square of MILES.
Linguistic Distance	A measure of the difficulty of learning a foreign language for English-speaking Americans.
CONC	A measure of minority language concentration, given by the percentage of the population aged eighteen to sixty-four in the region in which the individual lives, who reports the same non-English language as the individual.
COLONY*	Country of origin is a current or former colony of the US or the UK (excludes Puerto Rico).
REFUGEE*	Constructed using information on age, age at arrival, and birthplace, to identify the major sources of post-WWII refugees to the US.

Note: * = dichotomous variable.

III. THE 2000 US CENSUS DATA

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.[13] 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. The analyses are performed overall and separately for immigrants from Mexico and all other countries in Section IV. 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 of illegal aliens than migrants

[13] 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. Persons born in US Territories, such as Puerto Rico, are also excluded.

from other countries. They may be of special interest for these reasons. Further analyses of the immigrants from countries other than Mexico are presented in Section V, with this extension being characterised by the estimation of separate equations for immigrants with mother tongues close to English and for those with mother tongues distant from English.[14] The variables are defined briefly in Table 1, and in detail, and the means and standard deviations are reported, in Appendix A.

The 2000 US Census has three questions on language. The first of these is a streaming question: "Does this person speak a language other than English in the home?". Individuals responding YES to this initial question were asked two further questions. The first of these elicited the non-English language spoken in the home ("What is this language?"), with only one language being coded. The second was "How well does this person speak English?". Individuals were asked to indicate one of four levels of proficiency: very well; well; not well; not at all. Five categories of English use/proficiency may be formed using this information, namely: (i) speaks only English 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.

Two broad measures of English skills have been constructed from this information for use in empirical work. The first is a polychotomous (five categories) variable. The second is a binary measure, which is set equal to one for individuals categorized as "proficient in English", and is set to zero for those not "proficient in English". Proficiency in the analysis by Chiswick and Miller (2007a) was based on analysis of the impacts that the various English skills had on earnings, and covered the first three groups in the five-category variable described above. Other definitions could be considered, however. For example, where the emphasis is on "native-like" language outcomes, those who speak only English at home, or if a language other than English is spoken in the home, the individual speaks English "very well", might be categorized as proficient.[15] Analyses presented in Chiswick and Miller (2007a) suggest that the binary

---

[14] The 2000 US Census did not ask for mother tongue. The "mother tongue" used in this analysis has been inferred from the home language, and where English is spoken at home, the languages of the country of origin.

[15] Espenshade and Fu (1997) pool these two groups to form a single "very well" category.

and polychotomous approaches to modeling English language proficiency offer similar insights. Given the focus in the current paper, the statistical analysis uses a dichotomous dependent variable where those who speak only English at home, or speak English "very well", are distinguished from all other individuals.

IV. ESTIMATES OF MODEL OF ENGLISH PROFICIENCY

A. The Base Language Model

Table 2 lists estimates of the models of English proficiency for all males and separately for Mexican and non-Mexican males. These estimates have been obtained through probit analysis of the binary dependent variable. A parallel set of estimates for females are presented in Table 4. Variables for age by single year of age at migration were included in the estimating equation and are discussed subsequently.

(i) Analysis for Males

Because this is a probit equation and not an OLS equation, 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 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.

Table 2
Probit Estimates of Language Models, Adult Foreign Born Men, by Origin, 2000[(a)]

Variables	Total Sample	Immigrants from All Countries except Mexico	Immigrants from Mexico
Constant	-0.123 (1.95)	-0.167 (1.99)	0.106 (0.92)
Education	0.093 (75.77)	0.115 (65.25)	0.058 (31.00)
Years Since Migration (YSM)	0.024 (15.32)	0.024 (12.99)	0.021 (7.71)
YSM Squared/100	-0.023 (6.42)	-0.029 (6.60)	-0.026 (4.91)
Abroad	-0.159 (4.89)	-0.208 (4.82)	-0.096 (1.93)

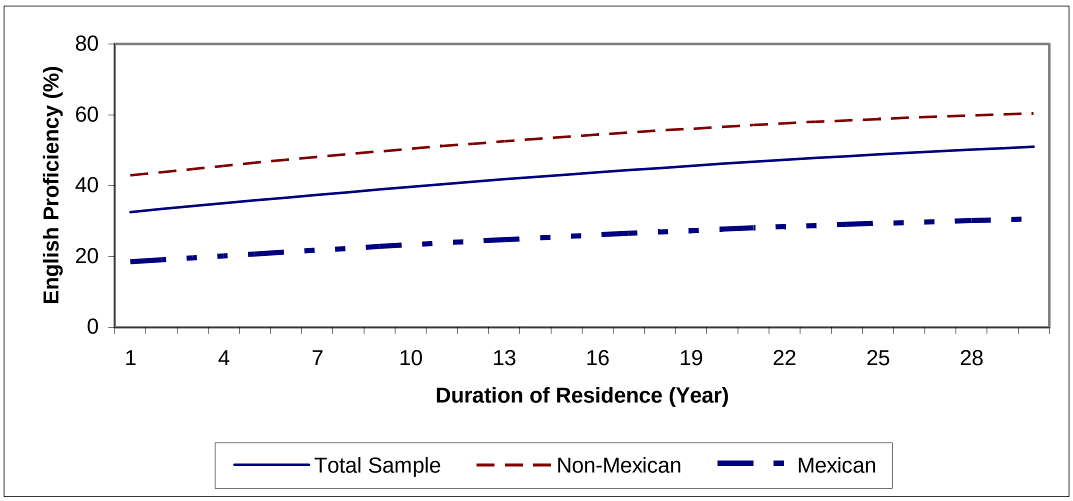
Married	0.016 (1.36)	0.024 (1.58)	0.027 (1.41)
With own children < 6 years only	-0.016 (0.99)	0.018 (0.85)	-0.064 (2.35)
With own children 6 to 17 years only	-0.063 (4.89)	-0.059 (3.71)	-0.053 (2.37)
With own children < 6 years <u>and</u> 6 to 17 years	-0.112 (7.05)	-0.058 (2.64)	-0.131 (5.45)
Non-Metropolitan	0.070 (1.47)	0.262 (3.03)	0.067 (1.16)
South	0.142 (12.84)	0.154 (10.76)	0.056 (2.97)
Miles (‘000) From Origin	0.228 (19.26)	0.178 (12.56)	-0.253 (1.98)
Miles (‘000) From Origin Squared	-0.026 (20.91)	-0.024 (17.18)	0.080 (1.46)
Linguistic Distance	-1.231 (26.88)	-1.290 (27.40)	(c)
Minority Language Concentration	-0.016 (21.36)	-0.019 (14.19)	-0.010 (6.04)
Colony	0.772 (42.54)	0.786 (41.85)	(c)
Refugee	-0.157 (6.04)	-0.102 (3.76)	(c)
Age at Migration(b)	Included	Included	Included
Sample Size	85,865	54,001	31,864
Prediction Success (%)	76.63	75.78	78.71
Pseudo R^2			
χ^2	28364.57	18578.21	4246.14

Notes: (a) ‘t’ statistics in parentheses, (b) For results, see Figures 3, 4 and 5, (c) variable not entered.
Source: 2000 US Census, 1% Public Use Microdata Sample.

Years of education are associated with higher levels of English proficiency, for the total sample and for each of the separate birthplace groups considered. The partial effect of a year of education is 3.7 percentage points greater proficiency for the total sample, 4.6 percentage points for immigrants from countries other than Mexico, and a more modest 1.8 percentage points for immigrants from Mexico. This complementarity between formal education and English proficiency is consistent with findings previously reported in the economics (*e.g.*, Chiswick and Miller (1995)(2007a)), sociology (*e.g.*, Espenshade and Fu (1997)) and linguistics (*e.g.*, Hakuta, Bialystock and Wiley (2003)) literatures.

Years of residence in the US are associated with greater proficiency in English, at least up to 40 to 50 years, which covers the vast majority of the sample. Moreover, in these analyses of data from the 2000 US Census, English proficiency continues to improve well after the 10 years of residence used as a cut-off in Bialystok and Hakuta (1999) in order to focus on ultimate attainment. Figure 1 presents the predicted relationships between English proficiency and immigrants' duration of residence in the US for each birthplace group. These predicted profiles have been constructed so that they pass through the means of the data (*i.e.*, the mean English proficiency for each sample is predicted when the duration of residence is at the sample mean value).

Figure 1
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 2.

Related to the strong period of residence effects is the negative impact associated with immigrants who first came to stay in the US more than five years ago but who were abroad in 1995, five years before the census. The impact of this sojourner or to-and-from migration behavior on the probit index (-0.159 for the pooled sample) is the equivalent of eight years of residence in the US. It is also apparent that the effect of this variable is much stronger for immigrants from countries other than Mexico than it is for immigrants from Mexico. The weaker negative effect for Mexico may explain the high frequency of sojourner migration. Alternatively, Mexican immigrants may invest little in US-specific human capital (including English language skills) in anticipation of a high rate of to-and-fro migration.

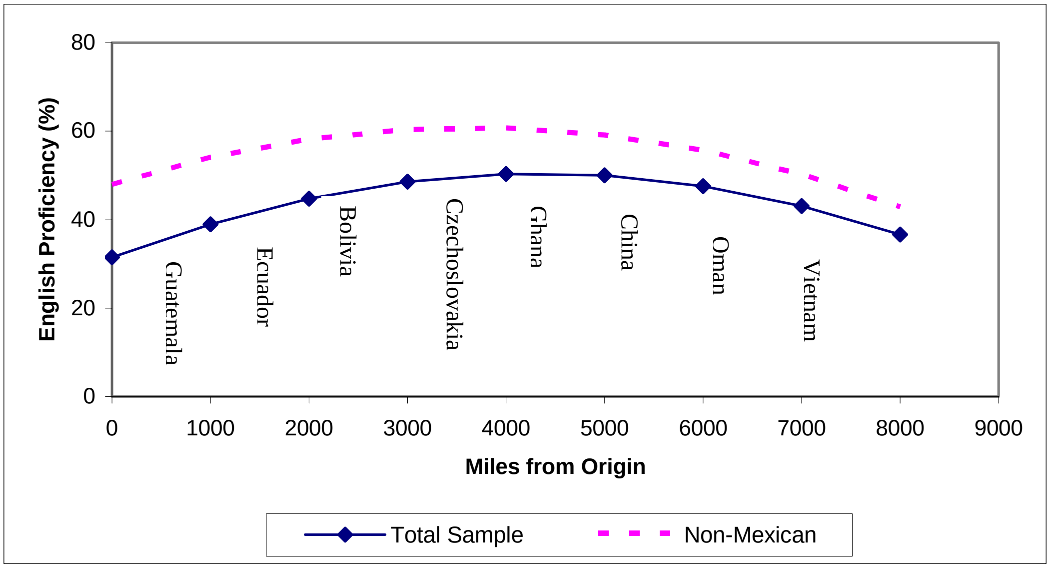
Among the family characteristics entered into the model (marital status, the age and number of children), only the variables for the presence of older children, or children aged under 6 years <u>and</u> 6-17 years (*i.e.*, families where there must be at least two children present) are consistently statistically significant. Both variables are associated with lesser proficiency in English among the fathers.

The geographic distance variable is statistically significant, and is associated with greater proficiency in English, up to around 4,400 miles in the pooled sample analyses, and 3,700 miles among immigrants from countries other than Mexico. This is consistent with the hypothesis of distance being related to favorable selectivity and a lower probability of return migration. The predicted relationships between English proficiency and the distance of the country of origin from the US are presented in Figure 2 for males for the total sample and for immigrants from countries other than Mexico. To aid the reading of this Figure, a number of country names are included to show the approximate locations of countries in relation to the US.[16]

For male immigrants from Mexico, the geographic distance variable (distance from the current US State of residence to Tijuana or Ciudad Juarez, whichever is shorter) is negative but is at the margin of statistical significance. The negative sign is unexpected. The US census does not include information on State of origin in Mexico. It will be shown below that the anticipated statistically significant positive coefficient is found for women.

[16] Espenshade and Fu (1997) report a similar, albeit more pronounced, non-monotonic association between English skills and distance from the US: They find that, compared to immigrants from countries up to 2,124 miles from the US, those from countries 2,124-4,856 miles from the US have superior English-speaking ability, those 4,856-6,104 miles from the US have English skills that are not significantly different, whereas the English-speaking skills of those from countries more than 6,104 miles from the US were less than those of immigrants from countries closer to the US.

Figure 2
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 2.

The linguistic distance variable is based on data provided in a paper by Hart-Gonzalez and Lindemann (1993). This reports language scores (LS) for 43 languages for English-speaking Americans of average ability after set periods (16 weeks and 24 weeks) of foreign language training. The range is from a low score (harder to learn) of 1.00 for Japanese to a high score (easier to learn) of 3.00 for Africaana, Norwegian and Swedish. Relevant details are reported in Chiswick and Miller (2005, Table 1). These scores suggest a ranking of linguistic distance from English among the various languages: from Japanese as being the most distant, through to Africaana, Norwegian and Swedish as the least distant. The number of languages for which a score is generated was extended in Chiswick and Miller (2005) to nearly all of the languages coded in the 2000 Census, other than Native American (American Indian) languages. The linguistic distance (LD) is measured in this analysis as the inverse of the linguistic score (LS) in Chiswick and Miller (2005), that is, LD = 1/LS.

The expected negative coefficient is found on the linguistic distance variable in the Table 2 results. Indeed, linguistic distance is one of the strongest predictors in the model. This is illustrated by the predictions contained in Table 3.

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

Linguistic Score	Linguistic Distance	Illustrative Languages	Predicted English Proficiency (%)	
			Total Sample	Excluding Immigrants from Mexico
1.0	1.00	Korean, Japanese	21.52	30.50
1.5	0.67	Vietnamese, Arabic	35.27	46.81
2.0	0.50	Polish, Indonesia	43.14	55.37
2.5	0.40	Portuguese, Italian	48.01	60.41
3.0	0.33	Norwegian, Swedish	51.29	63.69

Note: All other variables evaluated at their means.
Source: Authors' calculations based on Table 2.

The minority language concentration variable is highly significant in each of the analyses presented in Table 2. Living in a region with high concentrations of others, whether immigrants or the native born, who speak the immigrant's mother tongue, is associated with reductions in English proficiency. Neighborhood factors are clearly important to immigrants' English speaking ability.

(ii) Analysis for Females

The findings described above for males carry over to the analyses conducted for females (Table 4). In general the signs of the estimated coefficients, as well as the statistical significance of variables, are the same for females as they are for males. In other words, the model of English proficiency is robust between the genders.

Table 4
Probit Estimates of Language Models, Adult Foreign Born Women, by Origin, 2000[(a)]

Variables	Total Sample	Immigrants from All Countries except Mexico	Immigrants from Mexico
Constant	0.212	0.032	0.874
	(3.25)	(0.38)	(6.74)
Education	0.087	0.104	0.052
	(67.72)	(60.88)	(24.54)
YSM	0.026	0.029	0.018
	(17.18)	(15.81)	(5.94)

YSM Squared/100	-0.026 (7.48)	-0.032 (7.49)	-0.018 (2.79)
Abroad	-0.097 (2.70)	-0.175 (3.98)	0.089 (1.42)
Married	-0.063 (5.77)	-0.079 (5.94)	-0.014 (0.72)
With own children < 6 years only	-0.082 (4.72)	-0.058 (2.83)	-0.119 (3.54)
With own children 6 to 17 years only	-0.122 (9.88)	-0.112 (7.70)	-0.125 (5.11)
With own children < 6 years <u>and</u> 6 to 17 years	-0.176 (10.88)	-0.140 (6.67)	-0.181 (6.54)
Non-Metropolitan	-0.014 (0.27)	0.326 (4.03)	-0.218 (2.92)
South	0.156 (13.98)	0.163 (11.97)	0.102 (4.75)
Miles (‘000) From Origin	0.192 (15.90)	0.159 (11.30)	0.191 (2.95)
Miles (‘000) From Origin Squared	-0.025 (19.47)	-0.023 (16.53)	(c)
Linguistic Distance	-1.127 (27.72)	-1.152 (27.77)	(c)
Minority Language Concentration	-0.021 (25.84)	-0.022 (17.20)	-0.020 (10.55)
Colony	0.746 (41.94)	0.755 (41.35)	(c)
Refugee	-0.198 (7.54)	-0.138 (5.07)	-0.658 (4.46)
Age at Migration[(b)]	Included	Included	Included
Sample Size	83,832	58,000	25,832
Prediction Success (%)	76.48	75.21	79.62
Pseudo R^2			
χ^2	27,806.34	19,632.57	4,451.32

Notes: (a) ‘t’ statistics in parentheses, (b) For results, see Figures 3, 4 and 5, (c) variable not entered.
Source: 2000 US Census, 1% Public Use Microdata Sample.

There is one set of variables, however, which appears to impact more on the English proficiency of females than on that of males.[17] This is the family structure

[17] Another difference appears in the analysis for Mexican immigrants, where the distance variable has the expected significant positive sign for women.

variables. Thus, marital status is statistically significant, indicating that married females have lower English proficiency than their non-married counterparts, whereas this variable was insignificant in the analyses for males. In addition, children have a stronger (more negative) effect on the English skills of females than on that of males. The negative effect for women increases with the number and age of their children (under age 18) living at home. Both of these results are predicted by the theoretical model outlined in Chiswick and Miller (1995)(2007a), as a consequence of the lower attachment to the labor force of females compared to males, attributable to family responsibilities, including caring for children.

B. Testing the Critical Period Hypothesis

The age-at-migration data were entered into the model in the form of 64 dichotomous variables for individual ages at migration. This is a very flexible approach to quantifying the relationship between English proficiency and age at migration. In the analyses presented here, the ages at migration of zero and one are used as the omitted group, and all other age groups are compared to this one.[18]

A number of striking patterns emerge in Figures 3 to 5, which graph the relationship between proficiency in speaking English, as measured in this study, and age at migration by gender and origin (Mexico, non-Mexico, total).

First, proficiency in English declines monotonically with age at migration. This is consistent with both the biological and cognitive hypotheses.

Second, within country of origin, the patterns are very similar for men and women.

Third, the variability in the proportion proficient in English across adjacent ages at migration is greater the older the age at migration. This is presumably due to the smaller sample sizes at older ages at migration (greater sampling variability), because few immigrants come in middle age or older. For example, among male immigrants in

[18] This is equivalent to testing whether the onset of the critical period commences at the age of two, which would be consistent with Lenneberg's (1967) work. Recognizing that others have argued for an earlier age at onset, the analyses were repeated using only age zero as the omitted category: this did not change the conclusion in any material way.

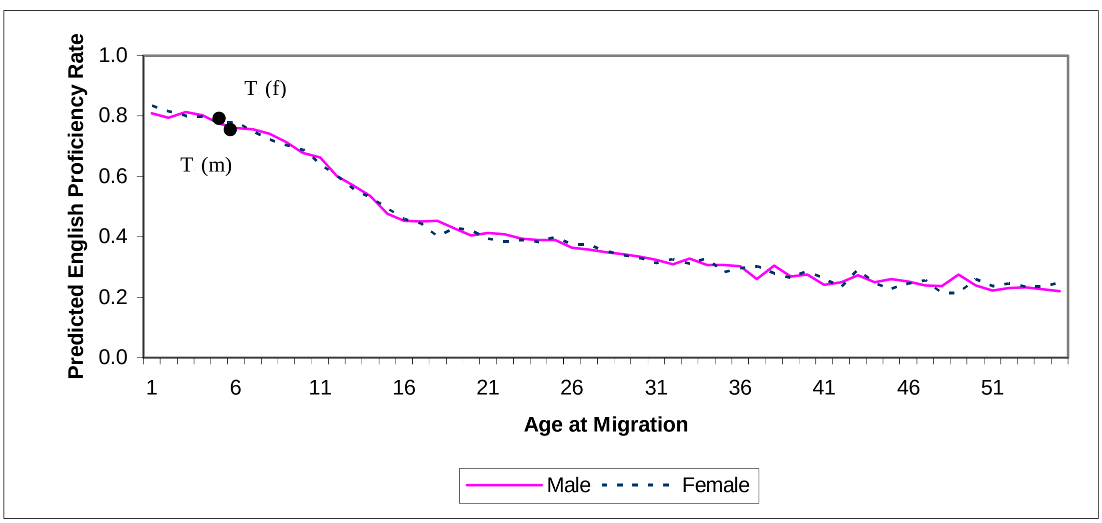
the sample under study, these were 3,829 immigrants who arrived at age 24, while there were only 255 who arrived at age 50.

Fourth, among non-Mexican immigrants, the decline in proficiency with age at migration is initially shallow (up to age 8), then exhibits a sharper decline that gradually diminishes, until the marginal effect of age at migration disappears after age 40. On the other hand, among Mexican immigrants there is a sharp decline in proficiency with age at migration until about age 16, after which there is little change. These patterns suggest important differences between immigrants from Mexico and all other countries.

It appears that if a "critical period" for language learning is defined as an age at which there is a sharp decline in the ability of immigrants to obtain proficiency in speaking English, no such critical period exists. The acquisition of English language proficiency (as defined above) declines with an increase in the age at migration. The pattern of decline is sharper for Mexican immigrants until about age at migration of 16 years, after which there is no important difference by age at migration. Among other immigrants, the decline is more gradual and disappears beyond an age at migration of 40.

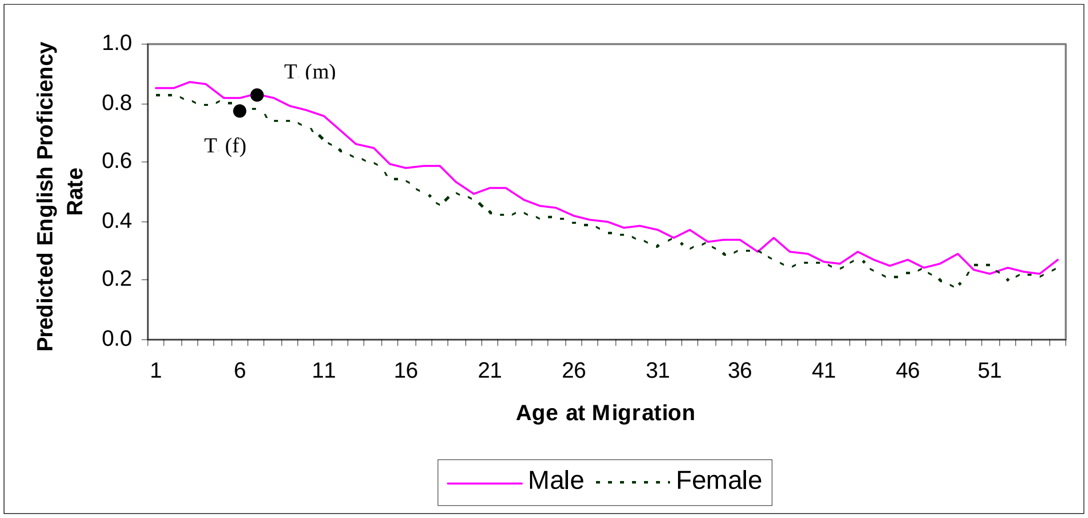
As a result, other variables the same, the proficiency rate among Mexican and other immigrants is about the same for those who immigrated at age zero or one (about 80 percent), lower thereafter for Mexican immigrants, until a parity is reached at about 30 percent at around an age at arrival of 40 and above.

Figure 3
Effects of Age at Migration on English Proficiency, Immigrants from non-English Speaking Countries, by Gender



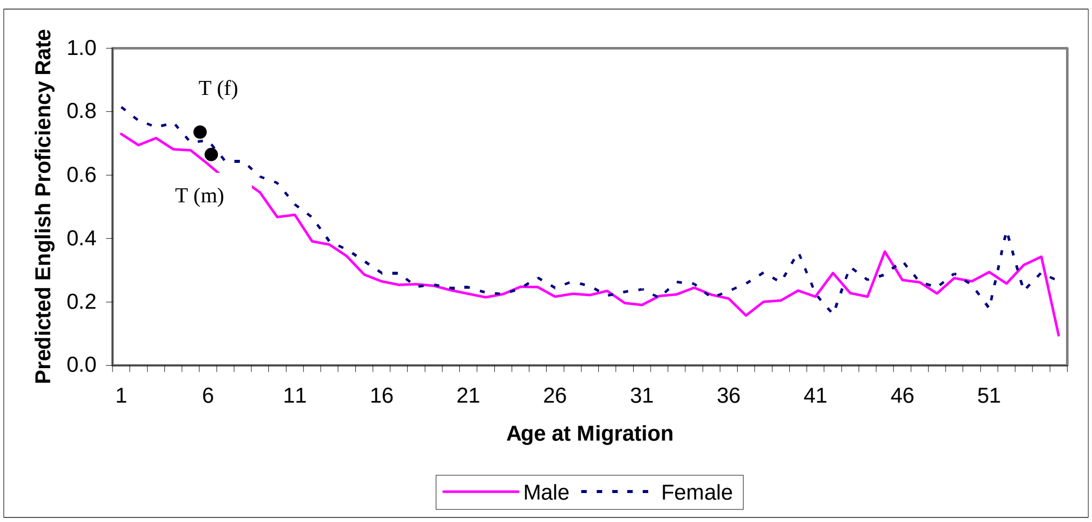
Source: Authors' calculations from 2000 US Census, 1% Public Use Microdata Sample.

Figure 4
Effects of Age at Migration on English Proficiency, Immigrants from Countries Other than Mexico, by Gender



Source: Authors' calculations from 2000 US Census, 1% Public Use Microdata Sample.

Figure 5
Effects of Age at Migration on English Proficiency, Immigrants from Mexico, by Gender



Source: Authors' calculations from 2000 US Census, 1% Public Use Microdata Sample.

Note that these analyses also show that there are ages at migration where the English proficiency of immigrants does not differ significantly from that of the benchmark group of very early arrivers, and that thresholds exist beyond which the age at migration effects are statistically significant.[19] These thresholds are denoted by "T" in the diagrams, with "m" and "f" in parentheses for males and females, respectively. For males these are age 6 for the total sample, age 9 for immigrants from countries other than Mexico, and age 6 for immigrants from Mexico. For females, the threshold ages are, in each instance, one year earlier than is the case for males. However, there are no adjacent age categories where the estimated effects in the language model differ from each other – which can be considered the purest interpretation of the critical period hypothesis.

V. ANALYSIS BY LINGUISTIC DISTANCE

In this section the data for immigrants from countries other than Mexico are disaggregated according to how difficult it might be for them to learn English. Models of English proficiency are estimated for each of the two samples derived.

The disaggregation by ease or difficulty of learning English is on the basis of the linguistic distance variable described above. A value of this that gave two, approximately equal-sized groups, was used to partition the sample.[20] Estimates are presented in Appendix B for the two groups. The relationships between age at migration and English proficiency for these groups are presented in Figures 6 and 7. The first of these graphs is for immigrants who have mother tongues distant from English and who, as a consequence, might be expected to find learning English more difficult. The second graph is for immigrants who have mother tongues close to English, and who therefore might be expected to find it easier to learn English.

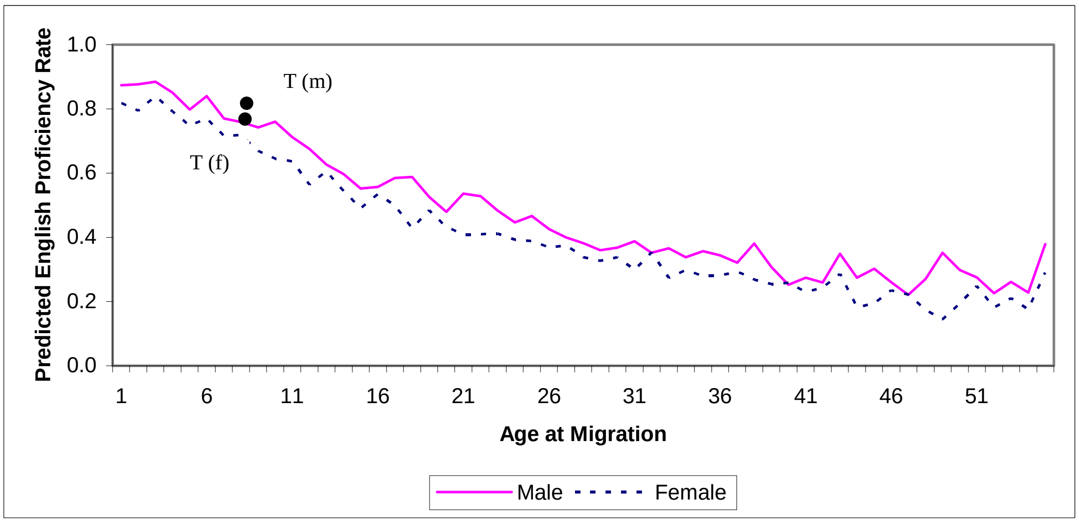
Figures 6 and 7 have the same general features as Figures 3 to 5. Thus, there is a negative relationship between English proficiency and age at migration for both immigrants with mother tongues close to English and for those with mother tongues distant from English. For immigrants with mother tongues distant from English, the

[19] Statistical significance refers to significant at the 5 percent level. Using a 1 percent level of significance lengthens the apparent period to a significant difference by only one year in most instances.

[20] The cut-off point uses a measure of linguistic distance of 0.5 (or linguistic score of 2), which is the score for Polish, Czech and Thai. These and languages with lower scores (Greek, Japanese, Hindi, Arabic, Mandarin) are in the group with mother tongues more distant from English. Languages with higher scores (Danish, Portuguese, Spanish, Italian) are in the group with mother tongues closer to English.

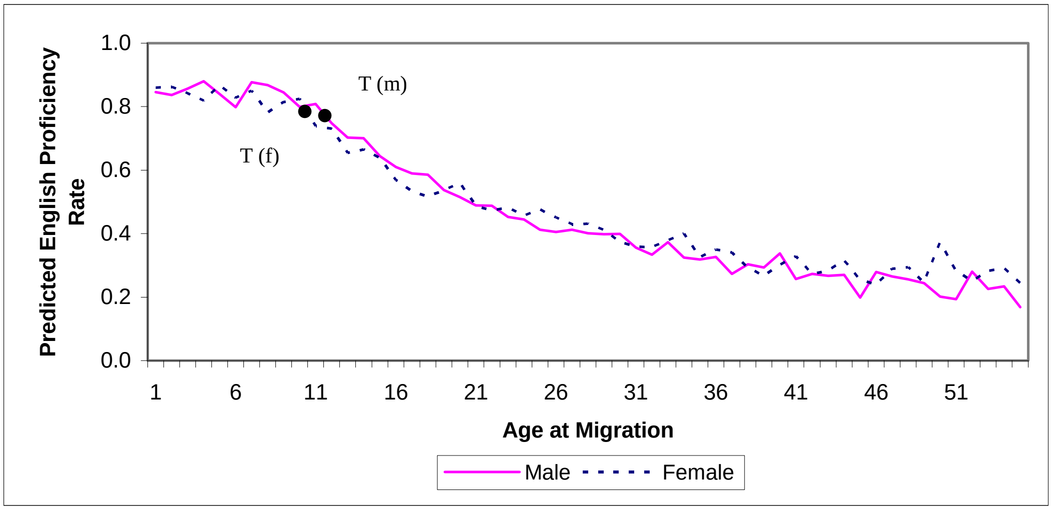
gradient of the relationship in Figure 6 is quite steep over the early ages at migration. For immigrants with mother tongues close to English, the decline in English proficiency with age at migration is initially modest, and it is therefore this group which generated the relatively flat profile of Figure 5. Thus, there is an interaction between linguistic distance and age at migration in an analysis of language learning. Age at migration matters more where it is more difficult to learn English.

Figure 6
Effects of Age at Migration on English Proficiency, Immigrants from Countries other than Mexico with Mother Tongues Distant from English, by Gender



Source: Authors' calculations from 2000 US Census, 1% Public Use Microdata Sample.

Figure 7
Effects of Age at Migration on English Proficiency, Immigrants from Countries other than Mexico with Mother Tongues Close to English, by Gender



Source: Authors' calculations from 2000 US Census, 1% Public Use Microdata Sample.

Note that the disaggregation by linguistic distance results in more variability in the predicted effects, owing to the smaller sample sizes used in the estimations. Nevertheless, the patterns that emerge are clear, and remarkably consistent for males and females.

VI. CONCLUSION

The analyses of the determinants of English proficiency among immigrants in the US from non-English speaking countries reported above are based on a model developed from an economics perspective that has clear parallels with research in sociology. This model has an emphasis on age at migration, which makes it very suitable for examining the critical period hypothesis.

Estimation of the model by gender for immigrants from Mexico, immigrants from other countries, and within the latter group, those with mother tongues close to English and those with mother tongues distant from English, shows clearly that proficiency in self-reported spoken English declines more-or-less monotonically with age at migration. This is consistent with both the biological and cognitive hypotheses on age-related differences in language acquisition.

However, the analyses could not identify an age at which there is a sharp decline in the ability of immigrants to acquire proficiency in speaking English. While the differences in English proficiency at ages of migration sufficiently far apart are statistically significant, the differences in English proficiency of adjacent age-at-migration categories do not differ significantly. These findings carry across to each of the sub-sample analyses. They are remarkably similar for males and females.

A critical period, defined with reference to a discontinuity, that is, a sharp, statistically significant, difference in English proficiency with age at migration, does not appear to exist for self-reported spoken English among immigrants in the US.

REFERENCES

Bialystok, Ellen, (1997). "The Structure of Age: In Search of Barriers to Second Language Acquisition", *Second Language Research*, Vol. 13, No. 2, pp.116-137.

Bialystok, Ellen and Hakuta, Kenji, (1999). "Confounded Age: Linguistic and Cognitive Factors in Age Differences for Second Language Acquisition", in D. Birdsong (ed.) S*econd Language Acquisition and the Critical Period Hypothesis*, Lawrence Erlbaum Associates, Mahwah, New Jersey.

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*, Vol. 3, No. 3, pp.243-268.

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

Chiswick, Barry R. and Miller, Paul W. (2003) "The Complementarity of Language and Other Human Capital: Immigrant Earnings in Canada", *Economics of Education Review*, Vol. 22, pp.469-480.

Chiswick, Barry R. and Miller, Paul W., (2005) "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 Miller, Paul W. (2007a). "Modelling Immigrants' Language Skills", *Research in Labor Economic* (forthcoming).

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

Cortes, Kalena E. (2004). 'Are Refugees Different from Economic Immigrants? Some Empirical Evidence on the Heterogeneity of Immigrant Groups in the United States', *Review of Economics and Statistics*, Vol. 86, No. 2, pp.465-480.

Dyen, Isidore, Kruskal, Joseph B. and Black, Paul, (1992). "An Indo-European Classification: A Lexicostatistical Experiment", *Transactions of the American Philosophical Society*, Vol. 82, No. 5, American Philosophical Society, Philadelphia.

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

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

Ginsburgh, Victor, Ortuño-Ortín, Ignacio and Weber, Shlomo, (2005). “Disenfranchisement in Linguistically Diverse Societies. The Case of the European Union”, mimeo, Departments of Economics, University of Alicante, Spain.

Hakuta, Kenji, Bialystok, Ellen and Wiley, Edward, (2003). “Critical Evidence: A Test of the Critical Period Hypothesis for Second-Language Acquisition”, *Psychological Science*, Vol. 14, No. 1, pp.31-38.

Hart-Gonzalez, Lucinda and Stephanie Lindemann, (1993). “Expected Achievement in Speaking Proficiency, 1993,” School of Language Studies, Foreign Services Institute, Department of State, mimeo.

Hyltenstam, Kenneth and Abrahamsson, Niclas, (2003). “Maturational Constraints in SLA”, pp.539-588 in Catherine J. Doughty and Michael H. Long (eds.) *The Handbook of Second Language Acquisition*, Blackwell Publishing, Oxford.

Krashen, S., (1973). “Lateralization, Language Learning and the Critical Period: Some New Evidence”, *Language and Learning*, Vol. 23, pp.63-74.

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

Lenneberg, Eric H., (1967). *Biological Foundations of Language*, John Wiley, New York.

Penfield, Wilder, and Roberts, Lamar, (1959). *Speech and Brain Mechanisms*, Princeton University Press, Princeton, New Jersey.

Scovel, Thomas, (2000). “A Critical Review of the Critical Period Research”, *Annual Review of Applied Linguistics*, Vol. 20, pp. 213-223.

Stevens, Gillian, (1992). “The Social and Demographic Context of Language Use in the United States”, *American Sociological Review*, Vol. 57, pp.171-185.

Stevens, Gillian, (2004). “Using Census Data to Test the Critical-Period Hypothesis for Second-Language Acquisition”, *Psychological Science*, Vol. 15(3), pp.215-216.

Wiley, Edward, Bialystok, Ellen and Hakuta, Kenji, (2004). "New Approaches to Using Census Data to Test the Critical-Period Hypothesis for Second-Language Acquisition", *Psychological Science*, Vol. 16(3), pp.341-343.

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 of the United States, 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 Variable:
English Language Fluency (LANG): LANG 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 "very well". The variable is set to zero where a language other than English is spoken in the home and the respondent speaks English either "well", "not well" or "not at all."

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 State 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 Chiswick and Miller, 2007a.) 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.

Location: The location variables record residence in a non-metropolitan area (NONMET) or in 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 came to the US to stay more than 5 years ago 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).

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. It takes a value of one for 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 Union (USSR) in 1950 or later, from China between 1950 and 1990, and Somalia or the former Yugoslavia between 1990 and 2000. All other immigrants are assigned a value of zero.

Linguistic Distance (LD): 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 (2005).

In the construction of this variable, foreign-born persons who speak only English at home are assigned the mean value of the linguistic score measure for individuals reporting a foreign language from their birthplace group. For the 3 percent of the sample who report a language for which there is no linguistic score, the sample mean value is assigned as the language score.

The variable in the regression equations is linguistic distance, which is one divided by the linguistic score, LD = 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*. For the analysis limited to immigrants from Mexico the distance from the capital in their state of residence in the US and Tijuana or Ciudad Juarez (whichever is shorter) is used.

Means and standard deviations for these variables are reported in Table A.1 (for men) and Table A.2 (for women).

Table A.1
Means and Standard Deviations of Variables in Language Models, Adult Foreign Born Men by Origin, 2000[(a)]

Variables	Total Sample	Immigrants from All Countries except Mexico	Immigrants from Mexico
Proficient in English (LANG)	0.427 (0.49)	0.532 (0.50)	0.254 (0.44)
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)
Years Since Migration (YSM)	16.135 (11.04)	16.413 (11.51)	15.634 (10.13)
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)
Linguistic Distance	0.508 (0.13)	0.543 (0.15)	(b)
Minority Language Concentration	10.447 (11.10)	4.798 (7.92)	20.645 (8.40)
Colony	0.150 (0.36)	0.233 (0.42)	(b)
Refugee	0.053 (0.22)	0.082 (0.27)	(b)

Sample Size	85865	54001	31864

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

Table A.2
Means and Standard Deviations of Variables in Language Models, Adult Foreign Born Women by Origin, 2000[(a)]

Variables	Total Sample	Immigrants from All Countries except Mexico	Immigrants from Mexico
Proficient in English (LANG)	0.431 (0.49)	0.510 (0.50)	0.254 (0.44)
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)
Years Since Migration (YSM)	16.573 (11.46)	16.762 (11.76)	16.127 (10.70)
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 <u>and</u> 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)
Linguistic Distance	0.520 (0.15)	0.552 (0.16)	(b)
Minority Language Concentration	9.860 (11.04)	4.803 (7.91)	21.778 (7.70)
Colony	0.160 (0.37)	0.228 (0.42)	(b)
Refugee	0.056 (0.23)	0.080 (0.27)	(b)
Sample Size	83832	58000	25832

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

APPENDIX B

Table B.1
Probit Estimates of Language Model, Adult Foreign Born Persons from All Countries except Mexico, by Gender and by Linguistic Distance, 2000[(a)]

	Males			Females		
Variables	Total Sample	Mother Tongue close to English	Mother Tongue distant from English	Total Sample	Mother Tongue close to English	Mother Tongue distant from English
Constant	-0.167 (1.99)	0.206 (1.15)	-1.019 (4.60)	0.032 (0.38)	0.068 (0.38)	-0.146 (0.69)
Education	0.115 (65.25)	0.103 (43.34)	0.126 (47.17)	0.104 (60.88)	0.102 (42.01)	0.110 (44.31)
Years Since Migration (YSM)	0.024 (12.99)	0.028 (10.40)	0.027 (9.51)	0.029 (15.81)	0.034 (13.22)	0.030 (11.03)
YSM Squared/100	-0.029 (6.60)	-0.035 (5.94)	-0.035 (4.99)	-0.032 (7.49)	-0.039 (6.85)	-0.039 (5.78)
Abroad	-0.208 (4.82)	-0.272 (4.59)	-0.171 (2.69)	-0.175 (3.98)	-0.111 (1.74)	-0.229 (3.71)
Married	0.024 (1.58)	0.023 (1.10)	0.011 (0.49)	-0.079 (5.94)	-0.084 (4.51)	-0.077 (4.02)
With own children < 6 years only	0.018 (0.85)	-0.044 (1.41)	0.063 (2.15)	-0.058 (2.83)	-0.085 (2.71)	-0.052 (1.86)
With own children 6 to 17 years only	-0.059 (3.71)	-0.079 (3.36)	-0.032 (1.45)	-0.112 (7.70)	-0.109 (5.03)	-0.123 (6.12)
With own children < 6 years and 6 to 17 years	-0.058 (2.64)	-0.120 (3.83)	-0.004 (0.14)	-0.140 (6.67)	-0.147 (4.80)	-0.144 (4.92)
Non-Metropolitan	0.262 (3.03)	0.521 (3.79)	0.044 (0.38)	0.326 (4.03)	0.340 (2.79)	0.290 (2.66)
South	0.154 (10.76)	0.121 (6.09)	0.130 (5.97)	0.163 (11.97)	0.147 (7.72)	0.159 (7.73)
Miles (‘000) From Origin	0.178 (12.56)	0.161 (8.59)	0.431 (6.70)	0.159 (11.30)	0.221 (11.97)	0.115 (1.85)
Miles (‘000) From Origin Squared	-0.024 (17.18)	-0.020 (8.56)	-0.043 (8.54)	-0.023 (16.53)	-0.030 (13.33)	-0.016 (3.33)
Linguistic Distance	-1.290 (27.40)	-1.951 (5.85)	-1.259 (19.82)	-1.152 (27.77)	-1.563 (4.82)	-0.934 (16.31)
Minority Language Concentration	-0.019 (14.19)	-0.016 (11.89)	-0.133 (11.49)	-0.022 (17.20)	-0.020 (14.82)	-0.064 (6.09)
Colony	0.786 (41.85)	1.229 (30.85)	0.650 (27.08)	0.755 (41.35)	1.250 (31.03)	0.654 (27.93)

Refugee	-0.102 (3.76)	-0.187 (3.89)	-0.067 (1.94)	-0.138 (5.07)	-0.184 (3.62)	-0.126 (3.78)
Age at Migration[(b)]	Included	Included	Included	Included	Included	Included
Sample Size	54,001	26,696	27,305	58,000	28,145	29,855
Prediction Success (%)	75.78	77.47	74.92	75.21	76.83	73.99
Pseudo R^2	0.254	0.289	0.236	0.245	0.277	0.223
χ^2	18578.21	10542.01	8676.09	19,632.57	10685.80	9284.67

Notes: (a) ‘t’ statistics in parentheses, (b) For results, see Figure 6 and 7.
Source: Authors’ calculations from 2000 US Census, 1% Public Use Microdata Sample.