DISCUSSION PAPER SERIES

IZA DP No. 449

# Do Enclaves Matter in Immigrant Adjustment?

Barry R. Chiswick
Paul W. Miller

March 2002

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

# Do Enclaves Matter in Immigrant Adjustment?

**Barry R. Chiswick**
*University of Illinois, Chicago and IZA, Bonn*

**Paul W. Miller**
*University of Western Australia*

Discussion Paper No. 449
March 2002

IZA

P.O. Box 7240
D-53072 Bonn
Germany

Tel.: +49-228-3894-0
Fax: +49-228-3894-210
Email: iza@iza.org

IZA Discussion Paper No. 449
March 2002

# ABSTRACT

## Do Enclaves Matter in Immigrant Adjustment?*

This paper is concerned with the determinants and consequences of immigrant/linguistic concentrations (enclaves). The reasons for the formation of these concentrations are discussed. Hypotheses are developed regarding "ethnic goods" and the effect of concentrations on the immigrant's language skills, as well as the effects on immigrant earnings of destination language skills and the linguistic concentration.
These hypotheses are tested using PUMS data from the 1990 U.S. Census on adult male immigrants from non-English speaking countries. Linguistic concentrations reduce the immigrant's own English language skills. Moreover, immigrant's earnings are lower the lower their English-language proficiency and the greater the linguistic concentration in their origin language of the area in which they live. The adverse effects on earnings of poor destination language skills and of immigrant concentrations exist independently of each other.

JEL Classification: J15, J24, J31, J61

Keywords: Immigrants, earnings, language skills, enclaves, ethnic/immigrant concentrations, ethnic goods, United States

Barry R. Chiswick
Department of Economics (MC 144)
College of Business Administration
601 S. Morgan Street (Room 2103 UH)
Chicago, IL 60607-7121
USA
Tel.: +1 (312) 996 2683
Fax: +1 (312) 996 3344
Email: brchis@uic.edu

* This paper was presented at the Conference on Magnet Societies, Loccum, Germany, June 2000, the Annual Meeting of the European Society for Population Economics, June 2000, Bonn, Germany, the Joint Center for Poverty Research, Northwestern University, February 2001, the Education and Employment Economics Group (UK) Annual Meeting, Leicester, July 2001, the Annual Meeting of the American Economic Association, Atlanta, January 2002 and Department of Economics seminars at George Washington University (November 2001) and Hebrew University, Jerusalem (December 2001). Comments received from participants at these presentations are appreciated although all errors of omission and commission are solely those of the authors. Miller acknowledges the support of an Australian Research Council grant. This paper received the Milken Institute Award for Distinguished Research in Regional and Demographic Studies, 2001.

## I. Introduction

This paper is concerned with the issue of immigrant/ethnic concentrations, that is, the tendency of immigrants to concentrate geographically by ethnicity or origin within the host country.[1] In particular, it is concerned with the consequences of enclaves or concentrations for two characteristics of immigrant adjustment – destination language proficiency and labor market earnings. Other aspects of immigrant life influenced by concentrations, including political participation, are beyond the scope of this paper.

There are two basic research questions of interest. One is the effect of immigrant concentrations on proficiency in destination language skills. The other is the direct effect of the immigrant's proficiency in the destination language and the effects of these immigrant concentrations on their labor market earnings. In particular, this study separates the direct effects and indirect effects via language proficiency of immigrant concentrations on earnings.

The application is to the United States. The methodology developed, however, could be applied to any immigrant receiving country for which there is appropriate census or survey data. Indeed, it is hoped that this paper will encourage estimations of similar models for other countries and other time periods, not merely to test the robustness of the approach but also to learn more about the specifics of immigrant adjustment in diverse settings.

---

[1] Earlier work on the determinants of immigrant concentrations include Bartel (1989) and Lazear (1999). Lazear (1999, p. S99) describes concentrations as forming "in large part because doing so enhances trade" in market and non-market goods and services.

Section II provides a brief introduction to the broader setting within which the issue of immigrant concentrations arises. Testable hypotheses are developed. Section III discusses the data used in the empirical analysis. In Section IV a model of dominant language acquisition is presented and estimated, with a particular focus on the effects of immigrant/linguistic concentrations on dominant language proficiency. Section V is the analysis of the earnings of immigrants with a particular focus on the effects of the immigrant's destination language skills and living in a linguistic concentration area on the respondent's labor market earnings. The paper closes (Section VI) with a summary and conclusion, with implications for public policy.

## II. Immigrant Concentrations: Hypotheses

### (A) Immigrant Flows

A characteristic of the late 20th century that is surely to continue into at least the early 21st century is an increase in the movement of people across international borders. International migration has increased into the traditional immigrant receiving countries, such as the United States, Canada and Australia. Yet, international migration into traditional countries of emigration has also become commonplace. Italy, Ireland, Germany and Japan, among others, are now experiencing large net in-migration, or where restricted by law, as in Japan, pressures for in-migration as evidenced by illegal flows.

These migration flows have, in part, been "East" to "West," that is, from the former Soviet Union and the Eastern block countries to the United States, Canada, Germany and Israel. More pronounced, however, are the migration flows from the "South" to the "North," more precisely, from less developed countries (LDC's) to highly developed economies (HDC's). Unprecedented immigration flows have been experienced from Latin America to North America, from Africa to Western Europe, and

from Asia to North America, Western Europe, Australia and Japan. (Chiswick and Hatton, forthcoming).

An important characteristic of these international migration flows is that the immigrants are "different" from the natives. As was true of the immigration flows from Southern and Eastern Europe to North America at the turn of the 20th century, the immigrants to the developed countries at the turn of the 21st century "sound" and "look" different.

In spite of the world becoming a smaller place with the ease (falling cost) of the transmission of information and ideas, and the falling cost of transportation for people and goods, and hence the "Westernization" of much of the world, new immigrants are frequently distinctive. Although distinctive clothing, especially for men, is less common than in the past, immigrants as a group frequently differ from natives as a group in appearance, religion, customs, belief-systems, language and other characteristics associated with ethnicity.

**(B) Immigrant Concentrations**

The immigrant groups typically have a spatial distribution in their host countries that differ sharply from that of the native born. For obvious reasons, new immigrants typically settle in areas based on three characteristics (Bartel 1989). The first is "ports" of entry, near seaports in the past, near airports in the current era. The second is where family and friends (co-ethnics) from earlier migrations have settled. Even if the location choice of the first settler from the ethnic group is purely random among a set of equally attractive locations in a destination country, once that first settler is established, future settlers are no longer indifferent among destination sites. The third is where the jobs are,

that is, where the immigrants are most able to gain employment that makes best use of their skills, or lack thereof. With the passage of time "ports of entry" and "family and friends" become less central in deciding where to live in the host country, and immigrants tend to disperse to some extent.

Some interpret the "family and friends" or chain migration effect on immigrant formations of concentrations as "clannishness." Yet to say it is clannishness is to beg the question as "clannishness" per se has no content. An alternative interpretation, however, is that settling in areas with others from the same origin provides for economies in communication, information, consumption and in the labor market.

Where new immigrants differ from the host population in terms of language skills, communication in all spheres of life are that much more difficult. These communication costs can be reduced if the host population were to learn the immigrant's language. Yet, it is not cost effective for a majority host population speaking the dominant language to learn the myriad of new languages that minority immigrants bring with them from various linguistic backgrounds.

These communication costs are reduced when immigrants learn the dominant language of the destination country. Yet, this learning can be costly and cannot be done instantaneously in the destination, thus, to varying degrees new immigrants tend to lack proficiency in the dominant language of the host economy, unless dominant language proficiency is a requirement for entry. Moreover, as with the production of other forms of human capital and of market goods and services, beyond some point, costs per unit of improved proficiency increase with a faster speed of language acquisition. Thus, the

optimal acquisition of dominant language proficiency among immigrants takes time and for some, full proficiency may never be obtained in their lifetime.

Finally, these communication costs for the immigrants can be reduced by living and/or working in a linguistic concentration area. Not all members of the group need dominant language proficiency, and the earlier arrivals and those more efficient in language acquisition are more likely to become proficient. They can serve as either direct or indirect translators for communication between the enclave and the host society. The demand for this specialized function increases with the size of the linguistic minority group and decreases as the members of the group learn the dominant language or as the native population learns the immigrant language.

Even aside from issues of language skills, immigrant/ethnic concentrations provide information networks that can be very valuable in social interaction, consumption and employment activities. Natives of an area have acquired location-specific human capital, which includes information obtained directly and indirectly through established networks. Not being connected to host country information networks when they arrive, immigrants have an incentive to create or "import" information networks through living in geographic concentrations with other new and longer term immigrants from the same origin.

**(C) Ethnic Goods**

Immigrants tend to differ from the native or host population in many dimensions related to ethnicity. They may differ in the foods they eat, the clothing they wear, the holidays they celebrate, the religion they practice, the media they read or hear (e.g., newspapers), their social organizations, and the languages they speak, among other

characteristics. There is frequently a tension between preserving the culture of the "old country" in the new setting and adopting the culture of the host country.

Let us call "ethnic goods" the consumption characteristics of an immigrant/ethnic group not shared with the host population, broadly defined to include market and non-market goods and services, including social interactions for themselves and their children with people of their same origin. To the extent that "ethnic goods" are important in the market basket and are distinctive, immigrants from a particular origin have a different market basket than immigrants from other origins and from the native born. The full cost of consumption of these ethnic goods varies with the price of purchased market goods and services and the value of time, but also with the importance and distinctiveness of the ethnic goods and the size of the group.[2]

There are certain fixed costs and economies of scale in the production and distribution of ethnic goods. Social interaction with others of the same origin (including finding an appropriate marriage partner) may involve little in the way of conventional market goods and services, but importantly involves the number of other individuals. The cost would decrease (presumably at a decreasing rate) the larger the size of the group. Up to a point, an ethnic religious institution (e.g., church, mosque, temple, or synagogue) has a lower per capita cost for members for the same type of facility providing the same level of services to the congregants if it is in a larger rather than in a smaller ethnic

[2] Distinctiveness is important as the ethnic goods of English immigrants to the United States would be much less distinctive than would those of, say, Chinese immigrants. To some extent the cost of ethnic goods can be reduced if the host society "adopts" the ethnic good, as, for example, often happens for certain foods, such as in Chinese restaurants. The "Americanized" version of the ethnic good may well differ from the version consumed in the origin or by members of the group in the destination.

community. There are fixed costs for buildings and hiring religious officials, among other items, including the probability that enough individuals will show up on a given occasion for the religious service.

The cost of "importing" into the community ethnic-specific goods (e.g., saris, Chinese vegetables, kosher meats) also varies with the size of the market because of economies of scale. Indeed, as the size of the community increases, the manner of "importation" may change from a family making a trip to a larger nearby community, to collective/cooperative efforts to place periodic bulk orders, to the establishment of a monopoly outlet, to many competitive outlets selling the product. The full price declines, the larger the size of the community.

The cost of living in an area then depends on the relative cost of ethnic goods, broadly defined, and the importance and distinctiveness of ethnic goods in the person's market basket. The cost of ethnic goods is lower, the larger is the size of the particular ethnic/immigrant community. The share of ethnic goods in the market basket is likely to be lower, the greater the extent of assimilation into the host society (that is, the closer culturally the group is in the origin to the host society), the longer the immigrant's duration of residence in the destination, and among the native born descendants of immigrants.

Ethnic goods have implications for living in an ethnic concentration area as well as for geographic differences in earnings. If ethnic goods, defined broadly, are an important part of the market basket, the person faces a higher real cost of living where ethnic goods are more expensive (an area where fewer co-ethnics live) than where they are less expensive (a high concentration area). Then the ethnic-immigrant would be

indifferent between a similar job in a high concentration area and a low concentration area only if the latter provided a higher nominal wage that was just sufficient to compensate for the higher cost of living.[3] Thus, ethnic goods can result in different geographic concentrations of various immigrant groups and differences in the pattern of regional wage differentials across immigrant groups and between immigrants and natives. The general observation would be lower nominal wages the larger the size of the concentration, other variables being the same. Note that the "ethnic goods" hypothesis regarding the negative relation between the concentration measure and earnings is an equilibrium situation based on differences in the real (ethnic-specific) cost of living. It reflects compensating wage differential.

When a new immigrant group is initially arrives in a destination it may be indifferent among alternative regions in the destination that are equally attractive in terms of job opportunities and ports of entry. The initial settlers would be immigrants with a lower demand for ethnic goods. Subsequent immigrants from this ethnic group will not be indifferent among the alternative destinations as ethnic goods will be cheaper where their co-ethnics have already settled. With the ethnic community established those with a higher demand for ethnic goods would find immigration that much more attractive.

[3] Workers of a given level of skill can be thought of as randomly drawing wage offers from a given distribution of wage offers available in the high concentration and the low concentration areas. If ethnic goods are an important part of their market basket, the ethnic-immigrants will move to or stay in a low concentration area only if their wage offer in this area exceeds by a sufficient margin the wage offer from the high concentration area to compensate for the higher cost of living. Once settled in a specific area explicit and implicit location-specific investments in human capital (in consumption and in the labor market) tend to reduce subsequent migrations. Thus, those who leave a high concentration enclave for a low or zero concentration area will tend to be those who receive a high wage offer in the latter location and those for whom ethnic goods (ties to the ethnic community) are least important.

New ethnic concentrations away from the original center in the destination can be formed under anyone of several scenarios. An individual with a very low demand for ethnic goods may settle elsewhere and gradually (or perhaps inadvertently) serve as a nucleus for others to follow. An individual with a high demand for ethnic goods may randomly receive a very high wage offer from the distribution of wage offers and settle in a new area. This person may serve as a nucleus and may even have an economic incentive to subsidize ethnic goods to encourage others to join him in the new location. Moreover, if a very "large" number of immigrants settle in the initial location and they are less than perfect substitutes in production for native workers, under the crowding hypothesis their wages decline relative to what they could earn in alternate locations with fewer (perhaps none) of their group. If the wage gap compensates for the higher cost of living of ethnic goods, a second enclave can be established. Thus, the number of enclaves or areas of concentration will vary systematically with the size of the immigrant/ethnic group and the distinctiveness and intensity of demand for ethnic goods.

**(D) Immigrant "Crowding"**

An alternative to the "ethnic goods" hypothesis is a labor supply or "crowding" hypothesis. If there are a large number of immigrants with a given skill level, and if they are not good substitutes in production for others with the same skill level, their earnings would be depressed. This is, however, a disequilibrium situation as immigrant workers with a given level of skill could receive higher real wages outside the enclave. The internal mobility of immigrant and native-born labor, and other factors of production, as

well as goods and services, would bring about factor price equalization, eliminating the negative relation between concentration and earnings.

The "crowding hypothesis" is not likely to be compelling for the United States. The U.S. has highly fluid labor, capital and product markets where inter-regional mobility is the norm rather than the exception. The largest single group of immigrants is from Mexico, and they tend to have low levels of skill, without a high degree of specialized skills.[4] As such, they are good substitutes in production for other low-skilled labor, whether native born or foreign born. Among the non-Mexican immigrants, the countries and languages of origin are numerous and skill levels are more highly dispersed. It is difficult to think of any groups in the U.S. that are sufficiently large and specialized with a low substitutability with native born and other foreign born workers. To the extent that a sudden exogenous infusion of immigrant labor with specialized skills impact a local labor market, disequilibrium earnings differentials would emerge, but would be dissipated over time with internal mobility of factors of production (including immigrant labor) and tradeable goods.

**(E) Consequences of Concentrations**

Limited destination language proficiency is likely to reduce the earnings potential of immigrants (Chiswick and Miller 1992, 1995). It raises the cost or lowers the efficiency of job search and in many jobs may restrict access (e.g., if there is a need to pass a test that requires proficiency) or merely lower productivity. There may also be discrimination in the labor market by the native population (either as employers, co-

[4] In the data under study for earnings, Mexican immigrants are 29 percent of the sample and have a mean schooling level of 7.9 years, in contrast to 13.3 years for the other immigrants.

workers or consumers) against those who are less proficient in the dominant language or who speak it with an accent. Working within a linguistic enclave is a mechanism for sheltering oneself from or mitigating the adverse labor market consequences of limited destination language proficiency.

Living and working within a linguistic concentration area has feedback effects on destination language proficiency. The greater the extent to which an individual can avoid communicating in the destination language, the slower is likely to be the rate of acquisition of dominant language skills. Consider two individuals: One lives in a large linguistic concentration area where one can work, consume, socialize and engage in other activities using the origin language. The other lives in a linguistically isolated area; communication can be done only in the dominant language. The latter may have a more difficult initial adjustment but has a stronger incentive to acquire destination language skills and has greater exposure that facilitates learning the destination language.

Thus, what has emerged in many developed countries is the existence of distinct immigrant communities that differ in language, culture and other characteristics from the host society. These immigrant/linguistic concentrations are expected to have an adverse effect on the immigrant's acquisition of dominant language skills. The immigrant's dominant language skills, as well as the size of the linguistic concentration area, will also affect the person's earnings, other things being the same. Greater proficiency would have a positive effect, and a larger concentration a negative effect on earnings. These hypotheses are tested in the empirical analysis.

## III. The Data

### (A) Defining the Population Under Study

The empirical analysis is performed using data from the 1990 Census of Population of the United States for adult male immigrants.[5] The U.S. Census provides a very large sample, a rich array of variables, and immigrants from diverse origins arriving at various periods of time. The analysis at this stage is limited to adult (non-aged) males as the analysis for females or aged males becomes more complex because of the need to model labor supply decisions, in addition to the language and earnings equations. Moreover, the formation of enclaves or concentrations is taken as exogenous for the individual in the empirical analysis, although there was a discussion in Section II as to why such concentrations are formed.

The data for the statistical analysis are from the five percent Public Use Microdata Sample from the 1990 Census. The sample is limited to males age 25 to 64 years who were foreign born but not from an English-speaking developed country. Thus, the native born, those born in a U.S. territory (e.g., Puerto Rico), born at sea or born abroad of American parents are excluded, as are those born in the United Kingdom, Ireland, Canada, Australia and New Zealand.

### (B) Defining the Variables

The English language proficiency variable comes from question number 15 in the census long form. Respondents were asked if there was a language other than English spoken at home (other than just a few words), and if so the identity of that language and

[5] The definition of the population under study and the variables used in the analysis are described in more detail in Appendix B.

how well they spoke English, where the response categories were Very Well, Well, Not Well and Not at All. For the purpose of this analysis, the foreign born who spoke only English or who spoke another language but reported that they spoke English "very well" or "well" were considered fluent; those who spoke English "not well" and "not at all" were considered not fluent.

The other dependent variable is earnings, which is the sum of wage, salary and self-employment income in 1989. Those with earnings of less than $100, including those with negative earnings, were assigned a value of $100. Those with zero weeks worked in 1989 were deleted from the sample for the analysis of earnings, as they were not labor market participants.

The enclave variable is a minority language concentration measure (CON). The 24 languages other than English most frequently spoken in the United States were identified. The speakers of these top 24 languages constitute around 94 percent of those reporting a foreign language spoken at home. For each of these 24 languages, for the 50 states and the District of Columbia, the percent of the states' population age 18 to 64 years (whether native or foreign born) speaking that language, was computed. The concentration measure for each respondent is the percent speaking the person's origin language in the state of current residence. For other languages, since the number of speakers is so low, the percent was assumed to be zero. Those who reported speaking only English were assigned the mean value of the concentration ratio for non-English language speakers in their birthplace group.

Within states, the density of population is less in rural areas than in urban areas. A variable for residence in a rural area (RURAL) is included because of a smaller

concentration of origin language speakers in rural than in urban areas. The other explanatory variables are straightforward and are discussed in Appendix B and as the variables are introduced in the text.

### (C) The Statistical Techniques

The main statistical methodology that is employed is ordinary least squares (OLS) with residuals corrected for heteroskedasticity, and where indicated below instrumental variables (IV) analysis.

## IV. Analysis of Language

### (A) The Language Model

This section presents the development of the model for dominant language proficiency. While largely based on previous work, in particular Chiswick and Miller (1995, 1998), the model is expanded to include new variables (refugees, persons from former colonies of English-speaking countries, and persons who lived abroad five years earlier). Particular attention is given to the variable measuring the degree of minority language concentration (CON).

The language proficiency model adopts a human capital approach in which destination language proficiency (LANG) is a function of three fundamental determinants, namely "exposure," "efficiency" and "economic incentives." Since the application is to the English language for immigrants in the United States 1990 Census, the discussion of these variables will be in this context. The principles apply to any destination language, country and data set.

### *(i)* Exposure Variables

"Exposure" refers to exposure to the destination language either pre- or post-immigration. The Census identifies country of birth but provides no other information on pre-immigration experiences relevant for acquiring English language proficiency. A set of country-of-origin dichotomous variables is included in the analysis to control for country of origin fixed effects. Western Europe (other than the UK and Ireland) is the benchmark. Moreover, a dichotomous variable is created for whether the origin was a colony (COLONY) of an English-speaking country, that is, of either the United States or the United Kingdom. Recall that respondents born in current U.S. territories are excluded from the analysis.

Post-immigration exposure to English can be measured in time units and in intensity per unit of time. Time in the destination is measured as the number of years since migration and its square (YSM, YSMSQ). It is expressed as a quadratic variable to allow for the effect of an extra year in the United States to be larger in the early years than in subsequent years.

The duration variable refers to when the immigrant first came to the United States to stay. Exposure to English in the United States may have been interrupted by sojourns outside the country after the initial migration. For immigrants in the United States for more than five years the variable "lived abroad five years ago" (ABROAD5) is unity if this was the situation, otherwise it is zero. It is expected that having lived outside the U.S. would be associated with lesser proficiency in English compared to otherwise similar immigrants who did not live elsewhere in 1985.

Intensity of exposure per unit of time in the United States can be measured by several variables. Of particular interest is the minority language concentration measure (CON) which is computed on a state level, as was discussed above. Within states the density of population is less in rural areas than in urban areas. A variable for residence in a rural area (RURAL) is included because of a smaller concentration of origin language speakers in rural than in urban areas within states.

For immigrants from Mexico the analysis also includes an index for Spanish language media, namely a variable for the number of radio stations in Spanish normalized for the size of the state in square miles and population (RADIO) (Chiswick and Miller 1998). Because of possible endogeniety in this variable, a predicted value (instrumental variables technique, IV) rather than an observed value for radio is used.

A marital status variable (MARR is unity if married, spouse present) is also included here. It is not possible in the 1990 Census to distinguish between pre- and post-migration marriages as it was possible to do in the 1980 Census (Chiswick and Miller, 1992).

***(ii)* Efficiency Variables**

"Efficiency" refers to the ability to convert exposure into language skills. Greater efficiency means more language skills are acquired for the same level of exposure. The efficiency variables include age at migration (age with years since migration held constant), years of schooling, whether the respondent may have been a refugee, and a measure of the "distance" between the origin language and English.

Older immigrants (AGE) at arrival have greater difficulty learning a new language. Age is entered as a quadratic variable (age and its square) as it is expected that

an extra year of age at migration would have a larger adverse effect among younger than among older immigrants.

Those with more schooling (EDUC) are assumed to be more able and to have more knowledge of the structure of languages, and hence are likely to be more efficient in learning English. It may also be that those with more schooling in the origin were exposed to English at higher grades prior to immigration, or that schooling in the U.S. enhanced proficiency.[6]

The refugee variable (REFUGEE) is included because refugees tend to be less favorably selected for a successful adjustment in the destination than are economic migrants. The migration decision of refugees is influenced to an important extent by factors other than the expectation of a successful adjustment. The refugee variable is based on country of birth and period of immigration.

Another efficiency variable is "linguistic distance" (DISTANCE), that is, a measure of how difficult it is for non-English speakers to learn English (see Chiswick and Miller, 1998). For example, Korean would be more "distant" from English than would be French. The more "distant" is the origin language from English, the lower the efficiency in learning English and hence the lower the expected proficiency in English.

[6] That higher levels of secular schooling is associated with greater proficiency in Hebrew among immigrants in Israel suggests that exposure to English in school prior to immigration is not the primary mechanism for the positive effect of schooling or English language skills in the United States (Chiswick and Ripetto, 2001).

### *(iii)* Economic Variables

"Economic incentives" is the most difficult conceptual variable to model. In principle, one would like to add an explanatory variable that measures the increment in expected earnings for a unit increase in proficiency for each respondent. Given currently available data it is not possible to do this. The economic benefits in the labor market and in other activities from increased proficiency in English would be greater the longer the expected duration in the United States. Immigrants from countries with a high propensity for return migration would expect a shorter period in the U.S. Data on emigration (EMIG) by country of origin are used for this purpose, but not for Mexico. The methodology for developing the country-specific emigration rates is not applicable to Mexico because of the very high proportion of illegal aliens in this group and the 1986 amnesty (Ahmed and Robinson 1994).

Immigrants from countries further from the U.S. are more likely to be favorably self-selected as they have higher costs of migration (Chiswick 1999). This implies a higher level of efficiency in learning English. They also have a lower return migration rate, again because of the higher migration costs. Those from origins a greater distance from the United States are therefore expected to be more fluent (Chiswick and Miller, 1998). Distance is measured as the distance in thousand of miles (XMILES) from the major city in the origin to New York, Miami or Los Angeles, whichever is the shortest. It is entered as a quadratic variable.

While state-specific (fixed) effects are not held constant because the concentration ratio is based on state data, a control variable is entered for Southern states (SOUTH).

**(B) Statistical Analysis**

The means and standard deviations of the language variable (LANG) and the explanatory variables, overall and separately for Mexican and non-Mexican immigrants, are reported in Appendix Table A1. The regression equations for English language proficiency are reported in Table 1 for all immigrants, non-Mexican immigrants, and Mexican immigrants.

The data are found to be consistent with the hypotheses developed above. In particular, English language proficiency is greater the higher the level of schooling, the longer the duration of residence (quadratic effect), the younger the age at immigration (negative effect of age), among those from a former British or American colony, and from countries more distant from the United States. It is less among refugees, among transients (i.e., immigrants who first came to the U.S. more than five years earlier – prior to 1985 – but who were outside the U.S. in 1985), and where the expected duration in the U.S. (emigration rate variable) is shorter. The linguistic distance variable is not statistically significant when country of origin fixed effects are included in the analysis, as is the case in Table 1, but it is significant with the expected sign when the country dichotomous variables are excluded from the equation. This arises from the close relation between country of origin and language of origin.

The minority language concentration variable (CON) is highly statistically significant as is the rural variable (RURAL), which is a proxy for the concentration of foreign language speakers within areas in states. According to the regression for all immigrants, going from a minority language concentration of zero to the mean value of 7.8 percent lowers the probability of being fluent in English by 3.1 percentage points,

which is 4.2 percent of the mean proficiency of 0.73 or 73 percent. Rural residence (5.5 percent of the foreign born) raises proficiency by 1.0 percentage point overall and by 2 percentage points among non-Mexican immigrants.

Among Mexican immigrants three variables reflect the effect of the linguistic concentration of Spanish speakers. One is the direct minority language concentration measure, the second is the rural variable, while the third is the (predicted) Spanish language radio station variable. The minority concentration measure and the radio station variable, but not the rural variable, are highly statistically significant with the expected negative signs.

Thus, the analysis of English language proficiency among immigrants from non-English origins in the United States indicates that the data are consistent with the model based on exposure, efficiency and economic variables. Moreover, it is found that linguistic concentrations or enclaves are associated with a lesser proficiency in English among all, Mexican and non-Mexican immigrants.

## V. Analysis of Earnings

### (A) The Earnings Model

The econometric analysis of earnings is based on the human capital earnings function, modified for immigrant adjustment (Chiswick 1978). In this specification, the natural logarithm of annual earnings (LNEARN) is regressed on years of schooling (EDUC), years of potential labor market experience and its square (EXP, EXPSQ), duration in the United States and its square (YSM, YSMSQ), the natural logarithm of weeks worked (LNWW), marital status (MARR), and place of residence (RURAL, SOUTH). Three dichotomous variables are added to the equation which take the value of

unity for immigrants whose race is Black, are Veterans of the U.S. Armed Forces, and who are Citizens of the United States. Two language variables are also added to this equation, the respondent's proficiency in English (LANG) which is unity for those fluent in English, as defined above, and zero otherwise, and the minority language concentration measure (CON).

**(B) Statistical Analysis**

The earnings equation is estimated separately for all immigrants, Mexican immigrants and non-Mexican immigrants. The means and standard deviations of the variables are reported in Appendix Table A-2, while Appendix Tables A-3, A-4 and A-5 report the regression equations for each group. A basic earnings function in these tables is reported in column (*i*) without the language and concentration variables, column (*ii*) adds the English language proficiency variable (LANG), column (*iii*) adds the concentration variable (CON) to the basic equation, column (*iv*) adds both variables, while column (*v*) substitutes a predicted English language proficiency variable obtained through the instrumental variables (IV) technique. (The auxiliary equation is reported in Appendix Table A-6.) A summary of the language and concentration variable results are presented in Table 2.

***(i)* Ordinary Least Squares Analysis**

As has been found elsewhere, the basic determinants of earnings among immigrants are also found to be important here (see Appendix Tables A-3 to A-5). For immigrants from non-English speaking countries, earnings increase with years of schooling (by about 5 percent per year of schooling), duration in the U.S. (at a decreasing rate), pre-immigration labor market experience (total experience when duration is held

constant), and weeks worked, and are higher for married men (by about 20 percent) and citizens (9 percent). Earnings are lower for immigrants who are veterans (8 percent) and those living in rural areas and in the south.

Patterns that are similar to the overall analysis are found when the analysis is done separately for Mexican and non-Mexican immigrants, although among Mexican immigrants, veteran status is associated with higher earnings (8 percent). Note that the effects of several variables reflecting human capital are smaller for Mexican immigrants than for other immigrants. This includes schooling, experience and weeks worked, but not duration in the United States.

The ordinary least squares analysis (OLS) in Appendix Tables A-3 to A-5 indicates that earnings are about 15 percent higher for all immigrants, Mexican immigrants and non-Mexican immigrants who are proficient in English, compared to those lacking proficiency. The difference is statistically significant and the magnitude of the effect and level of significance do not vary with whether the concentration measure is included in the analysis.

Assuming a long working life, the real rate of return on the investment in language proficiency can be estimated (approximately) as $r = b/k$, where r is the real rate of return, b is the regression coefficient of the language proficiency variable, and k is the number of full-year equivalents of lost earnings, including out of pocket expenditures, to go from not proficient ("not well," "not at all") to proficient (English only, "very well" or "well"). Then, if the coefficient of the language variable is $b = 0.15$ and if the cost is the equivalent of a full year potential earnings ($k = 1$), the rate of return is about 15 percent. If the cost were the equivalent of two years of full-time equivalent earnings ($k=2.0$), the

rate of return on the investment would be about 7.5 percent. If proficiency required the equivalent of only six-months foregone earnings (k=0.5), the estimated rate of return would be about 30 percent. The rate of return would be even higher if the positive effects of proficiency on weeks worked in the year were included in the calculation and if the consumption benefits from English language proficiency could be estimated. Thus, investments in English language skills appear to be profitable for immigrants from non-English speaking countries.

The concentration measure is also statistically significant in all three analyses. The coefficient and level of significance are also largely invariant with respect to the inclusion in the analysis of the respondent's fluency in English. Among all immigrants, going from a zero concentration area to the mean level (7.8 percent) lowers earnings by about 4.4 percent (that is, 7.8 times 0.0056 from Appendix Table A-3 column (*iv*)). For non-Mexican immigrants (mean concentration 3.9 percent) it lowers earnings by about 2.7 percent. Among Mexican immigrants, the mean of the concentration ratio is much higher (18.1 percent), but the coefficient of the concentration ratio is lower (-0.0033 compared to -0.0070 for other countries). For Mexican immigrants, the effect of going from a zero concentration to the mean concentration ratio is to lower earnings by about 6.0 percent. Thus, other variables the same, including the respondent's own proficiency in English, living in a linguistic/ethnic concentration area lowers the earnings of immigrants.[7]

[7] The labor supply or "crowding" hypothesis would imply a larger coefficient for Mexican immigrants than for the much more heterogeneous group of immigrants from other countries. That the opposite is found suggests that the negative relation between concentration and earnings is not a consequence of ethnic crowding in the labor market.

The effect of the concentration ratio on earnings varies systematically with the level of education. If an education – concentration ratio interaction term is added to the regression in Table A-3, column *iv*, it has a negative and highly significant effect.[8] That is, the adverse effect on earnings from living in a high concentration area is greater the higher the level of schooling. There is no effect for those with only five years of schooling, but the negative effect of living in a high concentration area grows larger at higher levels of schooling. Alternatively, this can be expressed as: the effect of education on earnings is smaller in the high concentration (enclave) area than in an area where fewer other individuals speak the same origin language.

***(ii)* Instrumental Variables Analysis**

There are several potential econometric problems with the ordinary least squares analysis using the respondent's reported level of English language proficiency. One problem is that language skills may be endogenous to, that is, determined by, earnings. Those who anticipate higher earnings if they were to become proficient will make greater investments to acquire proficiency (Chiswick and Miller 1995).

[8] Partial effects of education and the concentration ratio on earnings:

| | Table A-3 column *iv* | Table A-3, column *iv*, plus interaction |
|---|---|---|
| Education | 0.045 | 0.056 |
| | (82.9) | (78.3) |
| Minority Language Concentration | -0.0056 | 0.0062 |
| | (15.3) | (10.9) |
| Education-Concentration Interaction | ------ | -0.0012 |
| | | (26.9) |

A second problem is that there may be substantial measurement error in reported language skill. Purely random measurement error would bias the coefficient toward zero, but the measurement error need not be purely random. For example, those who are more successful in the labor market for unmeasured reasons may be more likely to overestimate their English language skills. A positive correlation in the measurement error terms could bias the coefficient upwards.

A third problem is that there may be dimensions of ability that are not in the equation but which enhance both English language proficiency and earnings. Those with greater innate ability among the foreign born may have superior English language skills and earn more, even though the higher earnings may be unrelated to their English proficiency. Yet there are no independent measures of ability in these data. This form of omitted variables bias would tend to overstate the true effect of language skills on earnings in an OLS equation.

Instrumental variables (IV) is a statistical technique that can correct for these potential problems by using a predicted rather than the observed value of language proficiency. An auxiliary regression is computed (Appendix Table A-6) which includes at least some variables that are not in the earnings function and which has a more complex functional form (various quadratic and interaction terms) to permit statistical identification. This auxiliary regression is used to obtain predicted values of the language variable, and it is these values, rather than the reported values, that are used in the earnings equation. Because the statistical identification is so dependent on variables that vary across countries of origin, a reliable instrumental variables model cannot be estimated using these data for immigrants from Mexico.

The results for the instrumental variables (IV) earnings function are reported in column (v) in Appendix Tables A-3 and A-4 and are summarized in Table 2 for all and non-Mexican immigrants. The IV technique results in a very large coefficient for the language proficiency variable. It implies about 80 percent higher earnings for those proficient in English in the all immigrant analysis.[9] Yet, similar very large coefficients on destination language skills have been found elsewhere and for other countries using this technique.[10] Perhaps the unbiased effect of English language fluency on earnings among immigrants is somewhere between the OLS and the IV estimates. Yet even the OLS estimate of about 15 percent implies a large pay-off from obtaining English language skills.

## VI. Summary and Conclusion

### (A) Summary

This paper has been concerned with whether immigrant linguistic concentrations or enclaves affect immigrant adjustment in terms of destination language proficiency and earnings.

The reasons for the development of these concentrations are discussed. New immigrants tend to settle near ports of entry, where previous immigrants from their origin ("friends and family") have settled and where their employment opportunities are best.

[9] The regression coefficient is Ln (1+X) = 0.59, where X is the percent increase in earnings. X is then 0.80 or 80 percent. Ln (1+X) is approximately equal to X when X is a small number. When Ln (1+X) = 0.15, X is approximately 16 percent.

[10] See Chiswick and Miller (1995), and the references therein, for the United States, Canada, Australia and Israel, and Dustmann and van Soest (1999) for Germany. The difference between the OLS and IV effects on earnings are much smaller in the United Kingdom (Dustmann and Fabbri 2000).

The "friends and family" or chain migration effect is a consequence of economies in communication, information, consumption, and the labor market.

"Ethnic goods" are market and non-market goods and services consumed by members of an immigrant/ethnic group that are not consumed by others. Ethnic-specific goods are an important factor in location choice. Because of economies of scale in the production of ethnic goods, the full cost of ethnic goods is lower the larger the size of the immigrant/ethnic group. Then an immigrant would be indifferent between working in two alternative areas only if the area with the high cost ethnic goods (lower concentration ratio) provided a higher nominal wage.

Several hypotheses emerge from the analysis. Linguistic concentrations are expected to have an adverse effect on the destination language proficiency of immigrants. Greater proficiency is expected to result in higher earnings and a larger linguistic concentration is expected to have a negative effect on nominal earnings.

The modeling of the language equation is based on three fundamental variables, exposure (pre- and post-immigration) to the destination language, efficiency in destination language acquisition, and economic incentives for destination language acquisition. Variables are developed to measure the effects of these concepts. The linguistic concentration ratio and the rural variable measure, in part, post-immigration exposure to the destination language.

The earnings equation is based on the standard human capital earnings function augmented for immigrant adjustment. Two additional variables are the immigrant's proficiency in the destination language and the minority language concentration ratio.

The empirical testing is done using adult (non-aged) male immigrants in the United States from non-English speaking countries, as reported in the 1990 Census 5 percent microdata sample. Immigrant language skills are found to vary positively with exposure to the destination language, efficiency in language acquisition and economic incentives. In particular, English language proficiency is greater the higher the level of schooling, the longer the duration of residence, the younger the age at immigration, the further the origin from the U.S., if the origin was a colony of the U.S. or the United Kingdom, if the immigrant was not a refugee, has a lower probability of return migration, and among longer term immigrants those who did not go back and forth between their native countries and the United States. A smaller minority language concentration ratio and living in a rural area, and hence living among a lower density of origin language speakers, are both associated with greater proficiency in English. Among immigrants from Mexico, greater access to Spanish language radio stations are associated with poorer English language skills.

Annual earnings are found to increase with skill level (schooling, experience, duration in the U.S.), and weeks worked, and are higher among married men, those living in urban areas outside the south, those who are citizens and those who are not black. Veteran status is associated with higher earnings among Mexican immigrants but lower earnings among other immigrants. In the Ordinary Least Squares (OLS) analysis earnings are higher by about 15 percent for those proficient in English, compared to those lacking fluency, and are lower for those living in an area with a higher minority language concentration ratio. The earnings advantage from proficiency is even greater when the

respondent's English language proficiency is estimated using the instrumental variables technique.

**(B) Policy Implications**

The answer to the question in the title is "yes." Enclaves matter for immigrant adjustment. Immigrant linguistic concentrations are associated with a lower level of proficiency in the destination language (English). Poorer English language skills result in lower nominal earnings. Living within a linguistic concentration area also results in lower nominal earnings, presumably because of the ethnic goods effect. Thus, linguistic concentrations have both an indirect effect (via destination language skills) and a direct effect on lowering the observed earnings of immigrants. The direct effect of concentration on earnings may be an equilibrium situation, where earnings differences reflect geographic differences in the cost of ethnic goods.

Immigrant/linguistic concentrations serve a useful role. They provide information networks and channels of communication in consumption and in the labor market for those without, or with only limited, destination-specific information and language proficiency, and they lower the cost of ethnic goods. On the other hand, they tend to retard the acquisition of or investment in destination-specific skills (e.g., language proficiency) and to lower nominal earnings. The assimilation or adjustment of immigrants is enhanced the smaller the extent of the concentration.

Incentives for immigrants to settle outside of concentrated areas for their group would be difficult, if not impossible, to implement. Focusing immigration on countries of origin "culturally similar" to the United States would be an unwarranted return to the pernicious national origins quota system in place from 1921 to 1965. A reduced

emphasis on family ties in issuing immigration visas, and placing a greater emphasis on the applicant's own skills is likely to increase the diversity of origins and reduce the extent of immigrant-linguistic concentrations.

Yet, in the highly mobile United States these concentrations tend to be first-generation, and at most also second-generation, phenomena. Reliance on self-correcting mechanisms, such as the acquisition of English language skills and the decline in the importance in the market basket of ethnic goods with a longer residence, is likely to be the most effective public policy.

## References


Ahmed, Bashir and J. Gregory Robinson (1994) Estimates of Emigration of the Foreign-Born Population, 1980-1990, Technical Working Paper No. 9, Washington, DC: U.S. Bureau of the Census.

Bartel, Ann (1989) "Where Do the New Immigrants Live?" Journal of Labor Economics, 7 (4), October, pp. 371-91.

Broadcasting and Cable Yearbook, 1994 (1994). New Providence, NJ: R.R. Bowker Publishers.

Chiswick, Barry R. (1978) "The Effect of Americanization on the Earnings of Foreign-Born Men," Journal of Political Economy, 86 (5), October, pp. 897-922.

Chiswick, Barry R. (1999) "Are Immigrants Favorably Self-Selected?" American Economic Review, 89 (2), May, pp. 181-185.

Chiswick, Barry R. and Timothy Hatton, (forthcoming), "International Migration and the Integration of Labor Markets, " in Michael D. Bordo, Alan M. Taylor and Jeffrey Williamson, eds. Globalization in Historical Perspective, Cambridge: National Bureau of Economic Research.

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

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

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

Chiswick, Barry R. and Gaston Repetto (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, 2001, pp. 204-228.

Dustmann, Christian and Francesca Fabbri (2000) "Language Proficiency and Labor Market Performance in the UK," University College London, London, Xerox.

Dustmann, Christian and Arthur van Soest (1998) "Language and the Earnings of Immigrants" Center for Economic Policy Research, London, Discussion Paper No. 2012.

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

Grimes, Joseph E. and Barbara F. Grimes, (1993) Ethnologue: Languages of the World, (13th ed.), Dallas: Summer Institute of Linguistics.

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

Lazear, Edward P. (1999) "Culture and Language" Journal of Political Economy, 107 (6, part 2), December, pp. S95-S126.

White, H. (1980) "A Heteroskedastic Consistent Covariance Matrix Estimator and a Direct Test for Heteroskasticity" Econometrica, Vol. 48, pp. 817-838.

**Table 1**
**Regression Estimates of Language Equation, Adult Foreign Born Men by Origin, 1990**

| Variable | Total Sample[a] | Excludes Immigrants from Mexico[a] | Immigrants from Mexico Only[b] |
|---|---|---|---|
| Constant | 0.409 | 0.478 | 0.440 |
| | (26.28) | (26.76) | (10.64) |
| Age | -0.010 | -0.011 | -0.006 |
| | (15.69) | (16.59) | (3.96) |
| Age Squared/100 | 0.003 | 0.006 | -0.003 |
| | (4.96) | (7.55) | (1.33) |
| Years of Education | 0.029 | 0.030 | 0.028 |
| | (141.10) | (119.57) | (64.67) |
| Years Since Migration (YSM) | 0.021 | 0.018 | 0.027 |
| | (100.30) | (83.11) | (48.35) |
| YSM Squared/100 | -0.025 | -0.023 | -0.027 |
| | (56.69) | (50.82) | (22.23) |
| Married | 0.033 | 0.020 | 0.053 |
| | (19.29) | (11.02) | (12.06) |
| Rural | 0.010 | 0.021 | 0.002 |
| | (3.00) | (6.93) | (0.26) |
| South | 0.013 | 0.010 | 0.028 |
| | (7.16) | (5.41) | (5.75) |
| S. Europe | -0.028 | -0.033 | (c) |
| | (9.47) | (10.88) | |
| E. Europe | -0.047 | -0.063 | (c) |
| | (12.19) | (15.41) | |
| USSR | -0.030 | -0.039 | (c) |
| | (4.75) | (6.19) | |
| IndoChina | -0.090 | -0.093 | (c) |
| | (9.19) | (9.22) | |
| Philippines | 0.024 | 0.014 | (c) |
| | (3.43) | (1.78) | |
| China | -0.123 | -0.128 | (c) |
| | (17.17) | (16.15) | |
| S. Asia | -0.011 | -0.020 | (c) |
| | (1.38) | (2.26) | |
| Other Asia | -0.036 | -0.044 | (c) |
| | (3.27) | (3.91) | |
| Korea | -0.202 | -0.207 | (c) |
| | (21.62) | (18.04) | |
| Japan | -0.108 | -0.116 | (c) |
| | (10.74) | (9.84) | |
| Middle East | 0.010 | 0.009 | (c) |
| | (2.52) | (1.10) | |
| Sub-Saharan Africa | 0.032 | 0.028 | (c) |
| | (6.28) | (2.79) | |
| Mexico | -0.067 | (c) | (c) |
| | (12.10) | | |
| Cuba | 0.044 | 0.040 | (c) |
| | (5.17) | (3.97) | |

| | | | |
|---|---|---|---|
| C. & S. America (Spanish) | -0.019 | -0.042 | (c) |
| | (4.10) | (8.25) | |
| C. & S. America (Non-Spanish) | 0.219 | 0.208 | (c) |
| | (32.97) | (30.76) | |
| Minority Language Concentration | -0.004 | -0.003 | -0.010 |
| | (17.81) | (9.73) | (7.84) |
| Linguistic Distance | -0.005 | -0.006 | (c) |
| | (0.36) | (0.44) | |
| Miles From Origin/1,000 | 0.050 | 0.054 | (c) |
| | (14.57) | (15.19) | |
| Square of Miles from Origin/1,000 | -0.034 | -0.038 | (c) |
| | (9.16) | (10.19) | |
| Refugee | -0.123 | -0.138 | (c) |
| | (32.12) | (35.19) | |
| Colony | 0.012 | 0.013 | (c) |
| | (3.53) | (3.90) | |
| Resident Overseas 5 Years Ago | -0.069 | -0.046 | -0.073 |
| | (11.34) | (6.76) | (3.13) |
| Emigration Rate | (c) | -0.010 | (c) |
| | | (2.37) | |
| Spanish Radio | (c) | (c) | -21.98 |
| | | | (4.11) |
| $\bar{R}^2$ | 0.3244 | 0.3176 | (d) |
| Sample Size | 237,766 | 169,253 | 68,512 |

Note: (a) = equation estimated using Ordinary Least Squares
(b) = equation estimated using Instrumental Variables (IV) estimator
(c) = variable not applicable
(d) R-squared not defined for the IV Model. IV estimator used for Spanish Radio variable.
't' statistics have been computed using White's (1980) heteroskedasticity-consistent covariance matrix estimator.

Source: 1990 Census of Population of the United States, Public Use Microdata Sample, 5 percent Sample.

**Table 2**

**Partial Effects on Earnings of the Language and Concentration Variables, Adult Foreign-Born Men from Non-English Speaking Countries, 1990**

| Variables | Total Sample | | Excludes Immigrants from Mexico | | Immigrants from Mexico Only | |
|---|---|---|---|---|---|---|
| | OLS | IV | OLS | IV | OLS | IV |
| Proficient in English | 0.148 | 0.592 | 0.151 | 0.678 | 0.146 | (a) |
| | (31.60) | (16.53) | (22.40) | (16.40) | (23.52) | |
| Minority Language Concentration | -0.0056 | -0.0039 | -0.0070 | -0.0050 | -0.0033 | (a) |
| | (15.25) | (9.62) | (11.77) | (7.85) | (7.13) | |

Note: Full regression equations reported in columns (iv) and column (v) of Tables A3 to A5. OLS refers to Ordinary Least Squares. IV refers to Instrumental Variables technique using predicted value of respondent's proficiency in English.

t-ratios in parenthesis

(a) IV equation not computed for Mexico

Source: Appendix Tables A3, A4 and A5.

**Appendix Table A1**
**Means and Standard Deviations of Variables, Sample Used for Language Model**

| Variable | Total Sample | Excludes Immigrants from Mexico | Immigrants from Mexico Only |
|---|---|---|---|
| English Proficiency | 0.730 | 0.808 | 0.524 |
| | (0.44) | (0.39) | (0.50) |
| Age | 37.79 | 40.92 | 36.83 |
| | (10.63) | (10.75) | (9.69) |
| Years of Education | 11.63 | 13.09 | 7.80 |
| | (4.99) | (4.27) | (4.69) |
| Years Since Migration | 15.21 | 15.43 | 14.64 |
| | (11.08) | (11.51) | (9.84) |
| Married | 0.655 | 0.673 | 0.610 |
| | (0.48) | (0.47) | (0.49) |
| Rural | 0.055 | 0.042 | 0.089 |
| | (0.23) | (0.20) | (0.28) |
| South | 0.237 | 0.234 | 0.244 |
| | (0.43) | (0.42) | (0.43) |
| S. Europe | 0.078 | 0.107 | (a) |
| | (0.27) | (0.31) | |
| E. Europe | 0.036 | 0.049 | (a) |
| | (0.19) | (0.22) | |
| USSR | 0.016 | 0.022 | (a) |
| | (0.12) | (0.15) | |
| IndoChina | 0.048 | 0.066 | (a) |
| | (0.21) | (0.25) | |
| Philippines | 0.051 | 0.070 | (a) |
| | (0.22) | (0.26) | |
| China | 0.062 | 0.085 | (a) |
| | (0.24) | (0.28) | |
| S. Asia | 0.049 | 0.067 | (a) |
| | (0.21) | (0.25) | |
| Other Asia | 0.012 | 0.016 | (a) |
| | (0.11) | (0.13) | |
| Korea | 0.031 | 0.044 | (a) |
| | (0.17) | (0.20) | |
| Japan | 0.015 | 0.020 | (a) |
| | (0.12) | (0.14) | |
| Middle East | 0.060 | 0.083 | (a) |
| | (0.24) | (0.28) | |
| Sub-Saharan Africa | 0.024 | 0.034 | (a) |
| | (0.15) | (0.18) | |
| Mexico | (0.276 | 0.00 | 1.00 |
| | (0.45) | (0.00) | (0.00) |
| Cuba | 0.051 | 0.070 | (a) |
| | (0.22) | (0.26) | |
| C. & S. America (Spanish) | 0.125 | 0.173 | (a) |
| | (0.33) | (0.38) | |
| C. & S. America (non-Spanish) | 0.009 | 0.012 | (a) |
| | (0.09) | (0.11) | |

| | | | |
|---|---|---|---|
| Minority Language Concentration | 7.784<br>(8.87) | 3.816<br>(6.19) | 18.178<br>(5.95) |
| Linguistic Distance | 0.515<br>(0.15) | 0.542<br>(0.17) | (a) |
| Miles From Origin | 3841.1<br>(2574.9) | 4756.6<br>(2475.4) | (a) |
| Refugee | 0.096<br>(0.29) | 0.133<br>(0.34) | (a) |
| Colony | 0.147<br>(0.35) | 0.203<br>(0.40) | (a) |
| Resident Overseas 5 Years Ago | 0.019<br>(0.14) | 0.017<br>(0.13) | 0.025<br>(0.16) |
| Emigration Rate | (a) | 2.049<br>(0.76) | (a) |
| Spanish Radio | (a) | (a) | 0.002<br>(0.01) |
| Sample Size | 237,766 | 169,253 | 68,512 |

Note: Standard errors in parentheses
(a) = variable not applicable.

Source: 1990 Census of Population of the United States, Public Use Microdata Sample, 5 percent Sample.

**Appendix Table A2**
**Means and Standard Deviations of Variables, Sample Used for Earnings Model**

| Variable | Total Sample | Excludes Immigrants from Mexico | Immigrants from Mexico Only |
|---|---|---|---|
| Natural Log of Earnings | 9.787 | 9.942 | 9.387 |
| | (1.03) | (1.04) | (0.90) |
| English Proficiency | 0.747 | 0.830 | 0.535 |
| | (0.43) | (0.38) | (0.50) |
| Labor Market Experience | 22.76 | 22.41 | 23.63 |
| | (11.46) | (11.53) | (11.25) |
| Years of Education | 11.79 | 13.30 | 7.90 |
| | (4.92) | (4.11) | (4.68) |
| Years Since Migration | 15.43 | 15.75 | 14.60 |
| | (10.85) | (11.30) | (9.52) |
| Married | 0.673 | 0.691 | 0.627 |
| | (0.47) | (0.46) | (0.48) |
| Rural | 0.057 | 0.044 | 0.091 |
| | (0.23) | (0.20) | (0.29) |
| South | 0.240 | 0.238 | 0.244 |
| | (0.43) | (0.43) | (0.43) |
| Race (Black) | 0.033 | 0.044 | 0.004 |
| | (0.18) | (0.21) | (0.06) |
| Citizen | 0.417 | 0.484 | 0.247 |
| | (0.49) | (0.50) | (0.43) |
| Veteran | 0.068 | 0.083 | 0.030 |
| | (0.25) | (0.28) | (0.17) |
| Log Weeks Worked | 3.752 | 3.774 | 3.693 |
| | (0.47) | (0.46) | (0.51) |
| S. Europe | 0.078 | 0.108 | (a) |
| | (0.27) | (0.31) | |
| E. Europe | 0.036 | 0.050 | (a) |
| | (0.19) | (0.22) | |
| USSR | 0.013 | 0.019 | (a) |
| | (0.12) | (0.14) | |
| IndoChina | 0.041 | 0.057 | (a) |
| | (0.20) | (0.23) | |
| Philippines | 0.053 | 0.073 | (a) |
| | (0.22) | (0.26) | |
| China | 0.061 | 0.085 | (a) |
| | (0.24) | (0.28) | |
| S. Asia | 0.051 | 0.071 | (a) |
| | (0.22) | (0.26) | |
| Other Asia | 0.011 | 0.016 | (a) |
| | (0.11) | (0.12) | |
| Korea | 0.031 | 0.043 | (a) |
| | (0.17) | (0.20) | |
| Japan | 0.015 | 0.021 | (a) |
| | (0.12) | (0.14) | |
| Middle East | 0.059 | 0.082 | (a) |
| | (0.24) | (0.27) | |

| | | | |
|---|---|---|---|
| Sub-Saharan Africa | 0.024 | 0.034 | (a) |
| | (0.15) | (0.18) | |
| Mexico | 0.279 | 0.00 | 1.00 |
| | (0.45) | (0.00) | (0.00) |
| Cuba | 0.051 | 0.069 | (a) |
| | (0.22) | (0.25) | |
| C. & S. America (Spanish) | 0.127 | 0.176 | (a) |
| | (0.33) | (0.38) | |
| C. & S. America (non-Spanish) | 0.009 | 0.012 | (a) |
| | (0.09) | (0.11) | |
| Minority Language Concentration | 7.834 | 3.850 | 18.129 |
| | (8.88) | (6.21) | (6.00) |
| Sample Size | 212,381 | 150,680 | 61,700 |

Note: Standard errors in parentheses
(a) = variable not applicable.

Source: 1990 Census of Population of the United States, Public Use Microdata Sample, 5 percent Sample.

**Appendix Table A3**
**Regression Estimates of Earnings Equation, Adult Foreign Born Men from Non-English Speaking Countries, 1990**

| | OLS | | | | IV |
|---|---|---|---|---|---|
| Variable | (i) | (ii) | (iii) | (iv) | (v) |
| Constant | 5.063 | 5.006 | 5.074 | 5.017 | 4.845 |
| | (173.18) | (171.67) | (173.47) | (171.96) | (150.58) |
| Years of Education | 0.049 | 0.045 | 0.048 | 0.045 | 0.035 |
| | (91.10) | (83.48) | (90.24) | (82.85) | (35.19) |
| Experience | 0.023 | 0.025 | 0.023 | 0.025 | 0.029 |
| | (35.72) | (38.04) | (35.77) | (38.04) | (38.55) |
| Experience Squared/100 | -0.037 | -0.038 | -0.038 | -0.038 | -0.041 |
| | (31.33) | (32.20) | (31.47) | (32.30) | (33.27) |
| Years Since Migration (YSM) | 0.028 | 0.025 | 0.028 | 0.025 | 0.017 |
| | (49.29) | (43.73) | (49.76) | (44.24) | (18.36) |
| YSM Squared/100 | -0.039 | -0.035 | -0.039 | -0.036 | -0.026 |
| | (30.24) | (27.56) | (30.54) | (27.88) | (16.85) |
| Log of Weeks Worked | 0.970 | 0.964 | 0.967 | 0.963 | 0.952 |
| | (135.52) | (134.94) | (135.43) | (134.88) | (131.21) |
| Married | 0.213 | 0.208 | 0.214 | 0.209 | 0.195 |
| | (55.22) | (54.02) | (55.43) | (54.23) | (47.54) |
| Rural | -0.037 | -0.038 | -0.043 | -0.044 | -0.047 |
| | (4.67) | (4.89) | (5.43) | (5.58) | (5.89) |
| South | -0.112 | -0.113 | -0.109 | -0.110 | -0.113 |
| | (26.11) | (26.36) | (25.40) | (25.71) | (25.90) |
| Race (Black) | -0.182 | -0.190 | -0.187 | -0.195 | -0.218 |
| | (12.36) | (12.95) | (12.68) | (13.22) | (14.48) |
| Veteran | -0.078 | -0.080 | -0.079 | -0.081 | -0.085 |
| | (10.25) | (10.48) | (10.39) | (10.61) | (11.12) |
| Citizen | 0.090 | 0.082 | 0.088 | 0.080 | 0.056 |
| | (21.36) | (19.35) | (20.87) | (18.94) | (11.83) |
| S. Europe | -0.063 | -0.060 | -0.058 | -0.056 | -0.049 |
| | (6.23) | (5.98) | (5.70) | (5.51) | (4.85) |
| E. Europe | -0.077 | -0.073 | -0.077 | -0.074 | -0.062 |
| | (6.40) | (6.09) | (6.44) | (6.13) | (5.14) |
| USSR | -0.133 | -0.125 | -0.134 | -0.127 | -0.103 |
| | (7.37) | (6.95) | (7.43) | (7.02) | (5.65) |
| IndoChina | -0.282 | -0.270 | -0.283 | -0.271 | -0.236 |
| | (23.21) | (22.31) | (23.31) | (22.42) | (19.02) |
| Philippines | -0.224 | -0.234 | -0.217 | -0.227 | -0.259 |
| | (21.11) | (22.07) | (20.39) | (21.39) | (23.42) |
| China | -0.274 | -0.254 | -0.270 | -0.251 | -0.193 |
| | (24.10) | (22.41) | (23.73) | (22.11) | (15.84) |
| S. Asia | -0.021 | -0.028 | -0.023 | -0.029 | -0.049 |
| | (1.83) | (2.41) | (2.00) | (2.55) | (4.13) |
| Other Asia | -0.201 | -0.203 | -0.202 | -0.203 | -0.208 |
| | (10.45) | (10.54) | (10.49) | (10.57) | (10.70) |
| Korea | -0.233 | -0.209 | -0.233 | -0.209 | -0.137 |
| | (14.95) | (13.41) | (14.94) | (13.43) | (8.25) |
| Japan | 0.347 | 0.357 | 0.347 | 0.357 | 0.389 |

| | | | | | |
|---|---|---|---|---|---|
| | (18.75) | (19.45) | (18.76) | (19.44) | (20.97) |
| Middle East | -0.098 | -0.104 | -0.099 | -0.105 | -0.122 |
| | (8.26) | (8.77) | (8.36) | (8.85) | (10.18) |
| Sub-Saharan Africa | -0.064 | -0.070 | -0.062 | -0.068 | -0.087 |
| | (3.38) | (3.71) | (3.29) | (3.62) | (4.54) |
| Mexico | -0.341 | -0.313 | -0.235 | -0.218 | -0.167 |
| | (37.39) | (34.39) | (21.28) | (19.80) | (14.23) |
| Cuba | -0.242 | -0.216 | -0.172 | -0.153 | -0.095 |
| | (21.54) | (19.22) | (14.35) | (12.77) | (7.44) |
| C. & S. America (Spanish) | -0.244 | -0.227 | -0.168 | -0.158 | -0.129 |
| | (25.62) | (23.89) | (15.93) | (15.05) | (11.96) |
| C. & S. America (Non-Spanish) | -0.081 | -0.100 | -0.073 | -0.092 | -0.150 |
| | (3.61) | (4.45) | (3.27) | (4.12) | (6.47) |
| Proficient in English | (a) | 0.151 | (a) | 0.148 | 0.592 |
| | | (32.26) | | (31.60) | (16.53) |
| Minority Language Concentration | (a) | (a) | -0.0062 | -0.0056 | -0.0039 |
| | | | (16.75) | (15.25) | (9.62) |
| $\bar{R}^2$ | .4157 | .4185 | 0.4164 | 0.4190 | (b) |
| Sample Size | 212,381 | 212,381 | 212,381 | 212,381 | 212,381 |

Note: (a) = variable not entered.
(b) R-squared not defined for the IV Model, IV estimator used for Proficient in English variable.
't' statistics have been computed using White's (1980) heteroskedasticity-consistent covariance matrix estimator.

Source: 1990 Census of Population of the United States, Public Use Microdata Sample, 5 percent Sample.

**Appendix Table A-4**
**Regression Estimates of Earnings Equation, Adult Foreign Born Men from Non-English Speaking Countries Other than Mexico, 1990**

| | OLS | | | | IV |
|---|---|---|---|---|---|
| Variable | (i) | (ii) | (iii) | (iv) | (v) |
| Constant | 4.824 | 4.757 | 4.839 | 4.773 | 4.542 |
| | (132.90) | (131.27) | (133.15) | (131.52) | (111.69) |
| Years of Education | 0.058 | 0.055 | 0.058 | 0.054 | 0.041 |
| | (85.04) | (77.84) | (83.96) | (76.98) | (33.96) |
| Experience | 0.023 | 0.025 | 0.024 | 0.025 | 0.030 |
| | (29.48) | (31.39) | (29.80) | (31.64) | (33.21) |
| Experience Squared/100 | -0.037 | -0.038 | -0.038 | -0.039 | -0.042 |
| | (24.66) | (25.42) | (25.05) | (25.77) | (26.90) |
| Years Since Migration (YSM) | 0.026 | 0.024 | 0.027 | 0.024 | 0.016 |
| | (37.82) | (33.87) | (38.05) | (34.15) | (15.63) |
| YSM Squared/100 | -0.037 | -0.034 | -0.038 | -0.035 | -0.024 |
| | (24.46) | (22.43) | (24.66) | (22.64) | (13.87) |
| Log of Weeks Worked | 0.994 | 0.991 | 0.994 | 0.990 | 0.978 |
| | (108.94) | (108.53) | (108.89) | (108.50) | (105.98) |
| Married | 0.218 | 0.215 | 0.217 | 0.215 | 0.205 |
| | (44.77) | (44.23) | (44.70) | (44.18) | (40.86) |
| Rural | -0.002 | -0.005 | -0.006 | -0.009 | -0.018 |
| | (0.18) | (0.43) | (0.50) | (0.72) | (1.48) |
| South | -0.087 | -0.088 | -0.091 | -0.091 | -0.093 |
| | (16.22) | (16.38) | (16.85) | (16.95) | (16.90) |
| Race (Black) | -0.189 | -0.197 | -0.196 | -0.203 | -0.228 |
| | (12.25) | (12.77) | (12.65) | (13.13) | (14.40) |
| Veteran | -0.093 | -0.095 | -0.094 | -0.095 | -0.101 |
| | (11.12) | (11.33) | (11.18) | (11.38) | (11.90) |
| Citizen | 0.107 | 0.098 | 0.105 | 0.097 | 0.069 |
| | (19.99) | (18.41) | (19.63) | (18.12) | (11.85) |
| S. Europe | -0.028 | -0.025 | -0.023 | -0.020 | -0.012 |
| | (2.72) | (2.46) | (2.18) | (1.96) | (1.17) |
| E. Europe | -0.069 | -0.063 | -0.071 | -0.065 | -0.045 |
| | (5.66) | (5.21) | (5.81) | (5.36) | (3.67) |
| USSR | -0.133 | -0.123 | -0.136 | -0.125 | -0.090 |
| | (7.28) | (6.74) | (7.44) | (6.90) | (4.84) |
| IndoChina | -0.266 | -0.251 | -0.269 | -0.253 | -0.199 |
| | (21.30) | (20.11) | (21.50) | (20.32) | (15.19) |
| Philippines | -0.225 | -0.232 | -0.217 | -0.225 | -0.253 |
| | (20.62) | (21.33) | (19.91) | (20.66) | (22.53) |
| China | -0.274 | -0.251 | -0.270 | -0.248 | -0.169 |
| | (23.67) | (21.67) | (23.33) | (21.39) | (12.94) |
| S. Asia | -0.041 | -0.044 | -0.043 | -0.047 | -0.058 |
| | (3.45) | (3.75) | (3.69) | (3.96) | (4.85) |
| Other Asia | -0.207 | -0.206 | -0.209 | -0.207 | -0.203 |
| | (10.68) | (10.63) | (10.77) | (10.71) | (10.36) |
| Korea | -0.237 | -0.208 | -0.238 | -0.210 | -0.112 |
| | (14.94) | (13.17) | (15.01) | (13.26) | (6.40) |
| Japan | 0.339 | 0.353 | 0.338 | 0.352 | -0.402 |

| | | | | | |
|---|---|---|---|---|---|
| | (18.12) | (18.99) | (18.07) | (18.93) | (21.06) |
| Middle East | -0.105 | -0.108 | -0.107 | -0.110 | -0.121 |
| | (8.72) | (9.01) | (8.89) | (9.16) | (9.99) |
| Sub-Saharan Africa | -0.071 | -0.074 | -0.068 | -0.071 | -0.082 |
| | (3.63) | (3.78) | (3.48) | (3.65) | (4.17) |
| Cuba | -0.230 | -0.203 | -0.141 | -0.120 | -0.051 |
| | (19.99) | (17.56) | (10.42) | (8.93) | (3.48) |
| C. & S. America (Spanish) | -0.217 | -0.197 | -0.124 | -0.112 | -0.067 |
| | (22.14) | (20.07) | (10.21) | (9.17) | (5.25) |
| C. & S. America (Non-Spanish) | -0.060 | -0.077 | -0.052 | -0.069 | -0.129 |
| | (2.62) | (3.36) | (2.25) | (3.00) | (5.44) |
| Proficient in English | (a) | 0.154 | (a) | 0.151 | 0.678 |
| | | (22.82) | | (22.40) | (16.40) |
| Minority Language Concentration | (a) | (a) | -0.0076 | -0.0070 | -0.0050 |
| | | | (12.71) | (11.77) | (7.85) |
| $\bar{R}^2$ | .3770 | .3792 | .3776 | .3797 | (b) |
| Sample Size | 150,680 | 150,680 | 150,680 | 150,680 | 150,680 |

Note: (a) = variable not entered.
(b) R-squared not defined for the IV Model. IV estimator for Proficient in English variable.

't' statistics have been computed using White's (1980) heteroskedasticity-consistent covariance matrix estimator.

Source: 1990 Census of Population of the United States, Public Use Microdata Sample, 5 percent Sample.

**Appendix Table A5**
**Regression Estimates of Earnings Equation, Adult Foreign Born Men from Mexico, 1990**

| | OLS | | | |
|---|---|---|---|---|
| Variable | (i) | (ii) | (iii) | (iv) |
| Constant | 5.208 | 5.194 | 5.279 | 5.254 |
| | (115.30) | (115.28) | (114.92) | (114.70) |
| Years of Education | 0.027 | 0.024 | 0.027 | 0.024 |
| | (29.94) | (26.67) | (29.77) | (26.58) |
| Experience | 0.015 | 0.016 | 0.015 | 0.016 |
| | (12.23) | (13.35) | (12.11) | (13.23) |
| Experience Squared/100 | -0.026 | -0.026 | -0.025 | -0.026 |
| | (12.74) | (12.90) | (12.64) | (12.81) |
| Years Since Migration (YSM) | 0.029 | 0.025 | 0.029 | 0.025 |
| | (29.47) | (25.36) | (29.82) | (25.70) |
| YSM Squared/100 | -0.037 | -0.033 | -0.037 | -0.033 |
| | (15.40) | (13.69) | (15.57) | (13.85) |
| Log of Weeks Worked | 0.918 | 0.913 | 0.918 | 0.913 |
| | (82.30) | (81.83) | (82.25) | (81.79) |
| Married | 0.207 | 0.199 | 0.208 | 0.200 |
| | (33.54) | (32.43) | (33.78) | (32.65) |
| Rural | -0.098 | -0.099 | -0.105 | -0.105 |
| | (10.58) | (10.73) | (11.27) | (11.29) |
| South | -0.184 | -0.184 | -0.174 | -0.175 |
| | (26.73) | (26.81) | (24.85) | (25.16) |
| Race (Black) | -0.039 | -0.055 | -0.038 | -0.054 |
| | (0.82) | (1.16) | (0.80) | (1.14) |
| Veteran | 0.087 | 0.078 | 0.085 | 0.076 |
| | (4.73) | (4.24) | (4.63) | (4.17) |
| Citizen | 0.042 | 0.028 | 0.040 | 0.026 |
| | (6.17) | (4.04) | (5.86) | (3.82) |
| Proficient in English | (a) | 0.149 | (a) | 0.146 |
| | | (23.98) | | (23.52) |
| Minority Language Concentration | (a) | (a) | -0.0039 | -0.0033 |
| | | | (8.53) | (7.13) |
| $\bar{R}^2$ | .4080 | .4135 | .4086 | .4139 |
| Sample Size | 61.700 | 61,700 | 61,700 | 61,700 |

Note: (a) = variable not entered.
't' statistics have been computed using White's (1980) heteroskedasticity-consistent covariance matrix estimator.

Source: 1990 Census of Population of the United States, Public Use Microdata Sample, 5 percent Sample.

**Appendix Table A6**
**Regression Estimates of Language Equation Used in IV Estimation,**
**Adult Foreign Born Men by Origin, 1990**

| Variable | Total Sample | Excludes Immigrants from Mexico |
|---|---|---|
| Constant | 0.350 | 0.398 |
| | (25.24) | (24.06) |
| Experience | -0.007 | -0.007 |
| | (23.10) | (21.09) |
| Experience Squared/100 | 0.004 | 0.004 |
| | (6.62) | (6.87) |
| Years of Education | 0.022 | 0.022 |
| | (67.15) | (65.15) |
| Years Since Migration (YSM) | 0.013 | 0.012 |
| | (54.17) | (44.62) |
| YSM Squared/100 | -0.019 | -0.016 |
| | (41.42) | (33.24) |
| Married | 0.028 | 0.016 |
| | (15.33) | (8.35) |
| Rural | 0.010 | 0.020 |
| | (3.07) | (6.62) |
| South | 0.005 | 0.004 |
| | (2.53) | (2.15) |
| Citizen | 0.064 | 0.055 |
| | (33.65) | (27.59) |
| Race (Black) | 0.068 | 0.065 |
| | (12.22) | (11.61) |
| Veteran | 0.019 | 0.012 |
| | (8.46) | (5.17) |
| Natural Logarithm of Weeks Worked | 0.026 | 0.022 |
| | (14.14) | (10.77) |
| S. Europe | -0.055 | -0.056 |
| | (18.47) | (18.16) |
| E. Europe | -0.072 | -0.087 |
| | (18.72) | (20.88) |
| USSR | -0.039 | -0.053 |
| | (6.22) | (8.22) |
| IndoChina | -0.156 | -0.134 |
| | (15.12) | (12.78) |
| Philippines | -0.065 | -0.038 |
| | (9.10) | (4.72) |
| China | -0.163 | -0.144 |
| | (21.67) | (17.56) |
| S. Asia | -0.102 | -0.072 |
| | (12.31) | (7.79) |
| Other Asia | -0.120 | -0.100 |
| | (10.44) | (8.15) |
| Korea | -0.242 | -0.196 |
| | (25.25) | (16.82) |
| Japan | -0.137 | -0.101 |

| | | |
|---|---|---|
| | (13.30) | (8.42) |
| Middle East | -0.038 | 0.008 |
| | (9.17) | (0.99) |
| Sub-Saharan Africa | -0.081 | -0.024 |
| | (10.71) | (2.04) |
| Mexico | -0.110 | (a) |
| | (18.91) | |
| Cuba | -0.024 | -0.008 |
| | (2.66) | (0.73) |
| C. & S. America (Spanish) | -0.057 | -0.053 |
| | (11.40) | (9.84) |
| C. & S. America (Non-Spanish) | 0.106 | 0.133 |
| | (13.69) | (15.99) |
| Minority Language Concentration (CON) | 0.028 | 0.022 |
| | (7.61) | (5.76) |
| Linguistic Distance | -0.002 | 0.009 |
| | (0.16) | (0.59) |
| Miles From Origin/1,000 | 0.035 | 0.047 |
| | (8.43) | (10.12) |
| Square of Miles from Origin/10m. | -0.012 | -0.023 |
| | (2.85) | (5.08) |
| Refugee | -0.116 | -0.113 |
| | (28.94) | (27.36) |
| Colony | 0.019 | 0.022 |
| | (5.70) | (6.57) |
| Resident Overseas 5 Years Ago | -0.066 | -0.045 |
| | (10.37) | (6.37) |
| Emigration Rate | (a) | -0.033 |
| | | (7.61) |
| CON * Years of Education/1000 | -0.024 | 0.103 |
| | (0.84) | (2.12) |
| CON * Experience/1000 | -0.206 | -0.270 |
| | (17.89 | (15.59) |
| CON * YSM/1000 | 0.527 | 0.666 |
| | (50.63) | (37.23) |
| CON * Linguistic Distance | -0.078 | -0.078 |
| | (9.26) | (9.11) |
| CON * Miles from Origin/1m. | 0.181 | -0.349 |
| | (1.40) | (2.41) |
| CON * Emigration Rate | (a) | 0.003 |
| | | (8.00) |
| $\bar{R}^2$ | 0.3345 | 0.3164 |
| Sample Size | 212,381 | 150,680 |

Note: (a) = variable note entered
't' statistics have been computed using White's (1980) heteroskedasticity-consistent covariance matrix estimator.

Source: 1990 Census of Population of the United States, Public Use Microdata Sample, 5 percent Sample.

## APPENDIX B

## DEFINITIONS OF VARIABLES

The variables used in the statistical analyses are defined below. Mnemonic names are also listed where relevant. The means and standard deviations are reported in Appendix Tables A1 and A2 for the samples used in the analyses of language attainment and earnings, respectively.

**Data Source:** 1990 Census of Population, Public Use Microdata Sample, 5 percent sample of the foreign born, except where noted otherwise.

**Definition of Population:** The sample used in this study comprises foreign-born men aged twenty-five to sixty-four, born in countries other than the English-speaking developed countries (UK, Ireland, Canada, Australia, New Zealand), territories of the United States, at sea or born abroad of American parents. Those with zero weeks worked in 1989 were deleted from the analysis of earnings, as they were not labor force participants.

**Dependent Variables:**

**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 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."

**Earnings** (LNEARN): The natural logarithm of the sum of wage or salary income and self-employment income (either non-farm or farm) received in 1989.

Individuals with earnings less than $100, including those with negative earnings, were assigned a value of $100.

**Explanatory Variables:**

**Minority Language Concentration** (CON): Each respondent is assigned a measure equal to the percentage of the population aged eighteen to sixty-four in the state in which he lives, who reports the same non-English language as the respondent. In the construction of this variable, only the twenty-four largest language groups nationwide are considered. In descending order there are: Spanish; French; German; Italian; Chinese; Tagalog; Polish; Korean; Vietnamese; Japanese; Portugese; Greek; Arabic; Hindi; Russian; Yiddish; Thai; Persian; French Creole; Armenian; Hebrew; Dutch; Hungarian; Mon-Khmer (Cambodian). These constitute 94 percent of all responses that 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 CON measure for other-language speakers of their birthplace group.

**Location:** The two location variables record residence of a rural area (RURAL) 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.

**Birthplace** (BIRTH): A number of non-English speaking birthplace regions are considered in the analyses: Western Europe; Southern Europe; Eastern Europe; former Soviet Union; Indochina; South Asia (which comprises the regions of British influence,

for example, India, Nepal, Pakistan); Other South-East Asia; Korea; Japan; Middle East and North Africa; SubSaharan Africa; Mexico; Cuba; Central and South America (Spanish influence); Central and South America (non-Spanish influence). The benchmark group (omitted category) in the regression analysis is Western Europe.

**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 U.K., U.S., Australia, New Zealand and South Africa are coded as colonies under this definition.

**Years Since Migration** (YSM). The categorical Census information on year of immigration is converted to a continuous measure using the following values: 1987-1990 (1.75 years); 1985-1986 (4.25 years); 1982-1984 (6.75 years); 1980-1981 (9.25 years); 1975-1979 (12.75 years); 1970-1974 (17.75 years); 1965-1969 (22.75 years); 1960-1965 (27.75 years); 1950-1959 (35.25 years); before 1950 (49.75 years).

**Lived Abroad Five Years Ago** (ABROAD5): This dichotomous variable is defined only for immigrants who have resided in the U.S. for more than 5 years. It is set equal to one if the individual lived abroad in 1985, otherwise it is set equal to zero for immigrants in the U.S. 5 or fewer years and for longer duration immigrants living in the U.S. in 1985.

**Radio** (RADIO): The number of radio stations broadcasting entirely or nearly entirely in Spanish in the state were obtained from *Broadcasting and Cable Yearbook, 1994* (1994), R.R. Bowker, New Providence, NJ, pp.B566-B567. In 1994, there were 315 Spanish language radio stations broadcasting in 25 states. Chiswick and Miller (1998) presents details. The number of Spanish-language radio stations in the state was

normalized by the area of the state to give the number of radio stations per thousand square miles. Then this variable was normalized by the number of Spanish speakers in the state of residence to give the number of Spanish language radio stations per unit of area per 10,000 Spanish speakers. This variable provides an index of the intensity of the infrastructure supporting the Spanish language in the state of residence. There were too few radio stations broadcasting in languages other than Spanish to compute a meaningful index for other languages. Because of the possible endogeneity of this variable, an instrumental variables (IV) approach was used.

**Marital Status** (MARR): 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 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 (2.5 years); completed fifth through eighth grade (7 years); completed ninth grade (9); completed tenth grade (10); completed 11th grade (11); completed 12th grade or high school (12); attended or completed college (14); Bachelor's degree (16); Master's degree (17.5); Professional degree (18); 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. Individuals who migrated from Cambodia, Laos or Vietnam in 1975 or later, Iran in 1980 or later, Cuba in 1960 or later, or the USSR and Baltic States 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 (1998, Appendix B). For the same number of weeks of instruction, a lower score (LS) represents less language facility, and, it is assumed, greater linguistic distance between English and the specific foreign language. For example, French is scored at 2.5 (in a range from 1 to 3), while Japanese is scored at 1.0. The language groups reported in the Hart-Gonzalez and Lindermann (1993) study are then matched to language codes in the 1990 Census using the Ethnologue Language Family Index published by Grimes and Grimes (1993). Adam Makkai, Professor of Linguistics, University of Illinois at Chicago, assisted in the matching of language codes, and in expanding the list of languages for which scores were assigned.

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.

**Emigration Rate** (EMIG): Yearly emigration rates of the foreign born by country of birth and sex are computed by dividing the yearly emigration levels between 1980-1990 from Ahmed and Robinson (1994) by the number of immigrants of the specific

birthplace-gender group in 1980 from the 1980 U.S. Census. Thirty-three countries are separately identified in the data, together with seven residual regions.

**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.

**Years of Experience** (EXP): This is computed as age minus years of education minus 5 (that is, EXP = AGE - EDUC - 5). A quadratic specification is used.

**Log of Weeks Worked** (LNWW): The number of weeks worked in 1989 is used in natural logarithmic form.

**Race:** This is a dichotomous variable, set to one if the individual is Black, and set to zero for all other racial groups (White, Asian and Pacific Islander groups, American Indian, other groups).

**Veteran Status** (VETSTAT): This is a dichotomous variable, set to one where the respondent is a veteran of the U.S. armed forces. In all other cases it is set to zero.

**Citizen** (CITIZEN): this is a dichotomous variable, set to one for individuals who are naturalized citizens.

# IZA Discussion Papers

| No. | Author(s) | Title | Area | Date |
|---|---|---|---|---|
| 434 | A. D. Kugler | From Severance Pay to Self-Insurance: Effects of Severance Payments Savings Accounts in Colombia | 4 | 02/02 |
| 435 | G. S. Epstein<br>M. E. Ward | Perceived Income, Promotion and Incentive Effects | 1 | 02/02 |
| 436 | A. Kunze | The Evolution of the Early Career Gender Wage Gap | 1 | 02/02 |
| 437 | M. Fertig | Evaluating Immigration Policy Potentials and Limitations | 6 | 02/02 |
| 438 | A. Voicu | Employment Dynamics in the Romanian Labor Market: A Markov Chain Monte Carlo Approach | 4 | 02/02 |
| 439 | G. Fella<br>P. Manzini<br>M. Mariotti | Does Divorce Law Matter? | 1 | 02/02 |
| 440 | G. Bertola<br>S. Hochguertel<br>W. Koeniger | Dealer Pricing of Consumer Credit | 7 | 02/02 |
| 441 | C. W. Sibley<br>P. P. Walsh | Earnings Inequality and Transition: A Regional Analysis of Poland | 4 | 02/02 |
| 442 | M. Lindahl | Estimating the Effect of Income on Health and Mortality Using Lottery Prizes as Exogenous Source of Variation in Income | 3 | 02/02 |
| 443 | K. L. Papps<br>J. O. Newell | Identifying Functional Labour Market Areas in New Zealand: A Reconnaissance Study Using Travel-to-Work Data | 1 | 02/02 |
| 444 | R. Lalive<br>J. Zweimüller | Benefit Entitlement and the Labor Market: Evidence from a Large-Scale Policy Change | 6 | 02/02 |
| 445 | G. S. Epstein | Informational Cascades and Decision to Migrate | 1 | 03/02 |
| 446 | E. Yashiv | Macroeconomic Policy Lessons of Labor Market Frictions | 6 | 03/02 |
| 447 | M. Fertig<br>C. M. Schmidt | Mobility within Europe – What do we (still not) know? | 6 | 03/02 |
| 448 | M. P. Keane<br>E. S. Prasad | Inequality, Transfers and Growth: New Evidence from the Economic Transition in Poland | 4 | 03/02 |
| 449 | B. R. Chiswick<br>P. W. Miller | Do Enclaves Matter in Immigrant Adjustment? | 1 | 03/02 |

An updated list of IZA Discussion Papers is available on the center‘s homepage *www.iza.org.*