Age at Arrival, English Proficiency, and Social Assimilation Among U.S. Immigrants

By Hoyt Bleakley and Aimee Chin*

June 2009

ABSTRACT

Are U.S. immigrants' English proficiency and social outcomes the result of their cultural preferences, or of more fundamental constraints? Using 2000 Census microdata, we relate immigrants' English proficiency, marriage, fertility and residential location variables to their age at arrival in the U.S., and in particular whether that age fell within the "critical period" of language acquisition. We interpret the differences between younger and older arrivers as effects of English-language skills and construct an instrumental variable for English-language skills. Two-stage-least-squares estimates suggest that English proficiency increases the likelihood of divorce and intermarriage. It decreases fertility and, for some, ethnic enclave residence. (JEL J24, J12, J13, J61)

* Bleakley: Graduate School of Business, University of Chicago, 5807 South Woodlawn Avenue, Chicago, IL, 60637, Center for Research and Analysis of Migration, and National Bureau of Economic Research (email: bleakley@chicagogsb.edu); Chin: Department of Economics, University of Houston, 204 McElhinney Hall, Houston, TX 77204-5019, and National Bureau of Economic Research (email: achin@uh.edu). We thank Chinhui Juhn, Marie Mora, Steve Trejo, two anonymous referees, and participants in the University of Washington Economics Seminar and May 2009 Workshop on the Economics of Immigration: Children of Immigrants and Temporary Migration for helpful comments and discussion. We also thank Mevlude Akbulut-Yuksel for excellent research assistance. Financial support from the National Institute of Child Health and Human Development (R03HD051562) is gratefully acknowledged. The authors bear sole responsibility for the content of this paper.

The increase in immigration to the United States in recent decades, much of it from non-English-speaking countries, has drawn attention to the role of English-language skills in immigrant assimilation.[1] There is evidence that English proficiency helps immigrants integrate economically into their new home—English proficiency raises wages, narrowing the wage gap between immigrants and U.S. natives. Less studied is whether sounding more "American" makes immigrants act more American as well. (Section I.A discusses the related literature.) This study addresses the connection between English proficiency and social integration.

The relationship between English proficiency and social assimilation among immigrants is a controversial topic in contemporary society. A commonplace hypothesis—often stated as fact—is that these variables are interrelated solely because of the culture or preferences of the immigrants themselves; that is, they choose not to integrate into U.S. society and they choose not to learn English. On the other side of the coin is the view that poor English proficiency is a constraint that impedes assimilation. In the present study, we examine this latter channel by quantifying the impact of English proficiency on marriage, fertility and residential location choices.

Understanding the effects of English proficiency on social outcomes also has important policy implications. Children of immigrants comprise a large and growing share of the U.S. population—in 2002, they made up 18.7% of the U.S. population under 18 (Randy Capps, Michael Fix and Jane Reardon-Anderson, 2003)—and their lower average education and earnings in adulthood have aroused concern.[2] Better knowledge about the family and neighborhood environment in which the children of immigrants grow up should improve our ability to design policies and programs to help them. Additionally, our ability to make

[1] The 2000 U.S. Census showed that 10.4 percent of the U.S. population is foreign born, up from 7.9 percent in 1990. Moreover, the 2000 U.S. Census also indicated that 47 million U.S. residents (age 5 and over) spoke a language other than English at home and 21 million spoke English less than fluently.
[2] Policymakers may be more interested in helping the children of immigrants than their parents because many of the children of immigrants are actually born in the U.S. or at least go through the U.S. school system.

demographic forecasts should improve; accurate population estimates are needed for a variety of policies, not just ones concerning immigrants.

A considerable challenge to estimating the effect of English proficiency on marriage, fertility and residential location outcomes is the endogeneity of English proficiency. English-language skills are correlated with many other variables that also affect these outcomes, such as ability and cultural attitudes. Additionally, reverse causality is possible. For example, immigrants who are married to U.S. natives may improve their English-language skills through interactions with their spouses, or immigrants who live in ethnic enclaves may develop worse English-language skills for the very reason that they live in enclaves. Thus, ordinary least squares regressions of social outcomes on English proficiency will mostly likely not estimate the causal effect.

Our strategy to identify the causal effect is based on a well-documented phenomenon from psychology: the critical period of language acquisition. Simply stated, young children learn languages more easily than older children and adults. We show in Section II that there is a strong association between immigrants' age at arrival in the U.S. and their English-language skills in adulthood using 2000 Census data on immigrants from non-English-speaking countries. (These data are described in Section I.C.) Indeed, the relationship we find is supportive of the critical period hypothesis: immigrants who arrive before age nine are uniformly fluent in English while those arriving later tend to have worse proficiency. Furthermore, we find no significant age-at-arrival effects on English proficiency for immigrants from countries where English is the dominant language. This is reassuring because, for these immigrants, age at arrival is decoupled from age at first exposure to English.

Age at arrival is also related to various social outcomes, especially for those immigrants from non-English-speaking countries. In Section III, we present these results in detail. For immigrants from Anglophone countries, marital status, various spousal characteristics (English proficiency among them), and fertility are, on average, all flat across age at arrival. In contrast, for immigrants from non-Anglophone countries, arriving later predicts higher marriage rates, higher fertility, and having spouses with lower human capital. Immigrants of either origin group are more likely to live in an "enclave" if they arrived later, which suggests a more general cultural (i.e., not just linguistic) mechanism at work, although the effect of age at arrival is a bit stronger for those from non-English-speaking countries.

We note a striking similarity in the age-at-arrival profile for English proficiency and the social outcomes. This fact suggests the following mechanism: childhood immigrants with first exposure to English after the critical period attain poorer English proficiency as adults, and their lower English proficiency in turn influences their socioeconomic outcomes. One complication with this interpretation, however, is that age at arrival probably affects immigrants' socioeconomic outcomes through channels other than language, such as through better knowledge of American culture and institutions. We therefore use immigrants from English-speaking countries to control for non-language-related effects of age at arrival. Again, arriving after the critical period predicts various social outcomes for immigrants of non-Anglophone origin, over and above that seen that seen for immigrants from Anglophone countries.

This analysis leads us to use an instrumental variable for English proficiency that is an *interaction* between immigrants' age at arrival and having a non-English-speaking country of birth. We implement this instrumental-variables strategy in Section IV. We start by considering marriage outcomes, and find that higher English proficiency decreases the probability of being married, both by decreasing the probability of ever having married and increasing the probability of being divorced. For those immigrants currently married with spouse present, we also examine spousal characteristics. We find that better English leads to more assimilation along several dimensions. Immigrants with greater English proficiency marry people who have better fluency in English, more education and higher earnings. Additionally, intermarriage (marrying outside one's nationality and ethnicity) is more likely. Next, we consider fertility outcomes, and find that immigrants with better English proficiency have fewer children though they are not significantly less likely to have a child. Finally, we consider residential location outcomes and find that immigrants with greater English proficiency are significantly less likely to live in "ethnic enclaves."

We then extend this analysis along several dimensions in Section V. We show that our main results are not sensitive to altering the specification or sample in ways that should increase the comparability between immigrants from English-speaking and non-English-speaking countries. Finally, we argue that language is, therefore an important constraint in immigrants' social outcomes, and offer other conclusions in Section VI.

I. Background and Data

A. Related literature

We are not aware of studies that address the problem of endogeneity of language skills when estimating the effect of language skills on marriage, fertility and residential location outcomes. However, a handful of studies estimate the correlation between language skills and these social outcomes. Studies with marriage outcomes include Gillian Stevens and Gray Swicegood (1987), Alberto Dávila and Marie T. Mora (2001), Xin Meng and Robert G. Gregory (2005) and Brian Duncan and Stephen Trejo (2007). Stevens and Swicegood find using data on U.S. immigrants that English proficiency raises the probability of intermarriage. Meng and Gregory find the same using Australian data, and furthermore find that intermarriage helps in earnings assimilation. Duncan and Trejo do not look at the effects of English proficiency *per se*, but instead examine the characteristics of Mexican Americans by whether they intermarried. They find that Mexican Americans who are married to non-Mexicans tend to speak English better, be more educated, be more likely to work and earn more compared to ones married to either Mexican immigrants or U.S.-born Mexicans. Similar differences prevail between the spouses of intermarried Mexican Americans and spouses of other Mexican Americans, consistent with assortative matching. Dávila and Mora examine marital status outcomes of recent Mexican immigrants, and find that the effect of English proficiency on the probability of being married varies by sex—women with low English proficiency are more likely to be married and the reverse holds for men.

Studies linking English proficiency to fertility outcomes include Ann Marie Sorenson (1988), and Swicegood, Frank D. Bean, Elizabeth H. Stephen and Wolfgang Opitz (1988). Swicegood et al. use 1980 Census data to estimate the effect of English proficiency on the fertility behavior of Mexican American women. They find that greater English proficiency is associated with a significantly lower number of children ever born and probability of having a child under three, especially among more educated women. Sorenson finds using Census data on Mexican American couples in Texas, New Mexico and Arizona that both wife's and husband's English proficiency reduce the probability of having an additional child at all parities. Although they do not consider English proficiency, Raquel Fernández and Alessandra Fogli (2006) correlate U.S.-born married women's fertility rates with lagged fertility rates in their country of ancestry, a relationship they characterize as cultural preferences.

Studies linking English proficiency to residential location outcomes include Edward Funkhouser and Fernando A. Ramos (1993) and Maude Toussaint-Comeau and Sherry L.W. Rhine (2004). Funkhouser and Ramos find that Dominican and Cuban immigrants with greater English proficiency are more likely to live outside of ethnic enclaves in the U.S. as opposed to either Puerto Rico or ethnic enclaves in the U.S. Toussaint-Comeau and Rhine find that less English proficient Hispanics are more likely to live in Chicago's Hispanic enclaves. Interestingly, Edward P. Lazear (2007) looks at the reverse relationship—the effect of living in an ethnic enclave on one's English proficiency—and finds a positive association. That researchers have used English proficiency as a dependent and explanatory variable underscores our point that the relationship between English-language skills and social outcomes is complex, and conventional ordinary least squares estimates are unlikely to provide the causal effect of English-language skills.

The main contributions of this study are (i) to note the pronounced differences in social outcomes as a function of age at arrival, especially for non-Anglophone-origin immigrants, and (ii) to use this relationship as a way to correct for endogeneity of English-language skills when estimating the effect of English-language skills on marriage, fertility and residential location outcomes. Another contribution is that we examine a broader set of marriage outcomes than has been considered by previous studies; in addition to marital status and intermarriage, we also consider spousal characteristics. As well, we estimate effects on a number of social outcomes using the same data and estimation framework; previously, each outcome was addressed in a separate study using varying samples and methodologies, making it more difficult to see the broader picture of the role of English in social assimilation.

B. Empirical strategy

To identify the causal effect of English-language skills on social outcomes, we take advantage of the phenomenon that younger children acquire language skills more easily than older children and adults. This window of easier language learning is known in psychology as the "critical period of language acquisition." It appears to be linked to physiological changes in the brain (Eric H. Lenneberg, 1967): maturational changes starting just before puberty reduce a child's ability to acquire second languages.[3] If exposure to the language begins during the critical period, acquisition of the language up to native-like proficiency is almost certain. If first exposure commences afterward, the individual's proficiency in that language is less assured.

These biological constraints in language acquisition generate a distinct relationship between immigrants' age at arrival in the U.S. and their English proficiency that varies by whether English was spoken in the origin country. In particular, for immigrants from non-English-speaking countries, first exposure to English begins when they arrive in the U.S. Thus, those who arrive at a younger age have an earlier age of first exposure to English and therefore a language-learning advantage compared to older arrivers. On the other hand, for immigrants from English-speaking countries, first exposure to English occurs early irrespective of age at arrival since English prevails in both the origin and destination countries.

Motivated by the similarity in shape between the age-at-arrival profile for English proficiency and the social outcomes (which we show below), we construct an instrumental variable for English-language skills based on age at arrival.[4] Age at arrival probably affects immigrants' socioeconomic outcomes through channels other than language, such as through better knowledge of American culture and institutions, thus it itself is unlikely to be a valid exclusion restriction. Incorporating immigrants from English-speaking countries into the analysis enables us to partial out the non-language effects of age at arrival. This is because, upon arrival in the U.S., immigrants originating from English-speaking countries encounter everything that immigrants from non-English-speaking countries encounter except a new language. Thus, any difference in child outcome between young and old arrivers from non-English-speaking

[3] Despite the evidence, including what we show below, there is some controversy surrounding the critical period hypothesis for language acquisition. (See Elissa L. Newport (2002) and David Birdsong (2006) for discussions of the theory and evidence.) For example, there has been debate over the specific mechanisms behind the negative relationship between age at first exposure to a language and language proficiency, when the critical period ends, and whether after the critical period the reduction in ability to acquire new languages declines sharply or gradually.

[4] This instrumental-variables strategy was used by Hoyt Bleakley and Aimee Chin (2004), who studied the effect of English-language proficiency on earnings.

countries that is over and above the difference from English-speaking countries can plausibly be attributed to language. In other words, the identifying instrument we use for English proficiency is an *interaction* between age at arrival and coming from a non-English-speaking country. The crucial assumption underlying the two-stage least squares estimates using such an instrument is that the non-language age-at-arrival effects are the same for immigrants from non-English and immigrants from English-speaking countries. (We relax this assumption in Section V.)

C. Data and descriptive statistics

We implement our empirical strategy using individual-level data from the 2000 U.S. Census of Population and Housing.[5] This is a large data set containing measures of English-language skills; a large number of observations is helpful for implementing any instrumental-variables strategy.[6] These measures are self-reported, and many researchers studying the relationship between language and earnings have used them.[7] Another attractive feature of the 2000 Census is that information is collected on all members of sampled households, which means individuals can be matched to co-resident spouses, enabling us to explore spousal characteristics as outcomes.

Our analysis is conducted using childhood immigrants currently aged 25 to 55.[8] We define a childhood immigrant as an immigrant who was under age 15 upon arrival in the U.S. For these immigrants, age at arrival is not a choice variable since they did not time their own

[5] Specifically, we combine the 1% and 5% samples from Integrated Public Use Microsample Series (IPUMS) (Steven Ruggles et al., 2004).

[6] The Census question based on which the English-ability measures in this paper are constructed is: "How well does this person speak English?" with the four possible responses "very well," "well," "not well" and "not at all." This question is only asked of individuals responding affirmatively to "Does this person speak a language other than English at home?" We have coded immigrants who do not answer "Yes" to speaking another language as speaking English "very well." We form an ordinal measure of English-speaking ability as follows: 0 = speaks English not at all, 1 = speaks English not well, 2 = speaks English well and 3 = speaks English very well.

[7] Robert Kominski (1989) reports that the Census measure of English-speaking ability is highly correlated with standardized tests of English-language skills and functional measures of English-language skills.

[8] For the purposes of this paper immigrant is defined as someone born outside the fifty states and the District of Columbia. This means that a person born in Puerto Rico is considered an immigrant, although legally he/she is a U.S. citizen at birth.

immigration but merely come with their parents to the U.S.[9] Given these criteria, individuals in the sample have spent a minimum of 11 years and a maximum of 55 years in the U.S., with an average of 30 years. Setting a reasonably high minimum age gives the immigrants enough time to learn English if they didn't know it already, and thus we expect differential time-in-country effects to be minimal. (Results are not sensitive to choosing a higher minimum time in country.) Setting a maximum time in country allows us to reducing selection biases arising from retirement or mortality.

We divide our sample into three mutually exclusive language categories: individuals from non-English-speaking countries of birth, countries of birth with English as an official language that have English as the predominant language, and other countries of birth with English as an official language.[10] The first category is our "treatment" group and the second is our "control" group. The last category is omitted from the main analysis, since we are not sure how much exposure to the English language immigrants from these countries would have had before immigrating. Table 1 provides the descriptive statistics for the treatment and control groups, with decompositions by age-at-arrival categories. Appendix Table 1 shows the decomposition of the sample by country of birth, and also presents our classification of countries by English-speaking status.

---

[9] According to the U.S. Citizenship and Immigration Services, immigrating parents may bring any unmarried children under age 21. We use a more restricted set of childhood immigrants: immigrants who were under 15 upon arrival (i.e., maximum age at arrival is 14). Using this lower age at arrival cutoff should mitigate the concern that many low-educated young men migrate on their own to the U.S. from Mexico and Central America to look for work, which makes age at arrival a choice variable and makes it less plausible that the non-language age-at-arrival effects estimated using immigrants from English-speaking countries apply to immigrants from non-English-speaking countries.

[10] We used *The World Almanac and Book of Facts, 1999*, to determine whether English was an official language of each country. Recent adult immigrants from the 1980 Census were used to provide empirical evidence of the prevalence of English in countries with English as an official language. English-speaking countries are defined as those countries from which more than half the recent adult immigrants did not speak a language other than English at home. The remaining countries with English as an official language are excluded from the main analysis. We made two exceptions to this procedure. First, despite the fact that Great Britain was not listed as having an official language, we included it in the list of English-speaking countries. Second, we classified Puerto Rico as non-English speaking even though English is an official language due to its colonial history.

II. Age at Arrival and English Proficiency

Figure 1, Panel A plots for each age at arrival the mean English-speaking ability among childhood immigrants from non-English-speaking countries and childhood immigrants from English-speaking countries. Immigrants from English-speaking countries are essentially all fluent in English, hence the flat age-at-arrival profile at "speaks very well", the highest level of proficiency measured by the Census. People who arrived at age nine or earlier from non-English-speaking countries speak English as well as their counterparts from English-speaking countries.[11] After age at arrival nine, people from non-English-speaking countries have significantly lower English-speaking proficiency, and indeed the disadvantage increases almost linearly with age at arrival.

These results are consistent with the critical period of language acquisition. Immigrants from non-English-speaking countries who arrive at older ages tend to have later ages of first exposure to the English language. For those arriving well within the critical period of language acquisition, a slightly later arrival does not depress English proficiency in the long run. Those who arrived as their critical period was coming to a close attained significantly worse eventual English skills. On the other hand, there is no relationship between age at arrival and English proficiency for immigrants from English-speaking countries—their first exposure to English is early regardless of when they arrived in the U.S.

Our empirical strategy compares younger and older arrivers from non-English-speaking countries after removing the age-at-arrival effects for immigrants from English-speaking countries. Thus in Figure 1, Panel B, we plot each age at arrival the mean English ability for immigrants from non-English-speaking countries with the mean for non-English-speaking countries subtracted out. In the case of English proficiency, there are no age-at-arrival effects for

[11] The higher English proficiency among early arrivers from non-English-speaking countries is an artifact of controlling for Hispanic status, a conventional demographic control variable. The curve is shifted down if the Hispanic dummy is excluded, but the shape of the curve is essentially unchanged.

immigrants from English-speaking countries; hence the graph of the difference in mean is the same as the graph of the mean for immigrants from non-English-speaking countries alone.

We can summarize the relationship between age at arrival and English-language skills in Figure 1, Panel B in a simple regression framework. In the analysis below, instead of estimating fifteen differences in means (for each age at arrival, 0 to 14), we estimate a parameterized difference that is allowed to vary by age at arrival. In particular, we impose the restriction that the difference is zero between childhood immigrants from non-English-speaking countries and childhood immigrants from English-speaking countries up through age at arrival nine, but has a linear relationship with age at arrival thereafter. This is a simple formulation that captures much of the co-movement between age at arrival and English-language skills displayed in Figure 1, Panel B. Symbolically, we use the following parameterization:

(1) $$k_{ija} = \max(0,a-9) \times I(j \text{ is a non-English-speaking country})$$

where a is age at arrival, I() is the indicator function, and j is country of birth.[12]

We estimate the relationship between English proficiency and age at arrival in the following equation:

(2) $$ENG_{ija} = \alpha_1 + \pi_1 k_{ija} + \delta_{1a} + \gamma_{1j} + \mathbf{w}_{ija}'\rho_1 + \varepsilon_{1ija}.$$

for individual i born in country j arriving in the U.S. at age a. ENG_{ija} is a measure of English-language skills, δ_{1a} is a set of age-at-arrival dummies, γ_{1j} is a set of country-of-birth dummies and $\boldsymbol{w}_{ija}$ is a vector of exogenous explanatory variables (e.g., age, sex, race).

The results from estimating Equation 2 are found in Table 2, Panel A. Because there are no endogenous variables on the right-hand side, Equation 2 can be consistently estimated using ordinary least squares (OLS). For each year past age nine that an immigrant from a non-English-speaking country arrives, the probability of speaking any English decreases 0.6 of a

[12] The results are unchanged when we parameterize the instrument differently. See Appendix Table 2.

percentage point (Row A1), speaking English well decreases three percentage points (Row A2) and speaking English very well decreases seven percentage points (Row A3). The ordinal measure of English-speaking ability, which encapsulates movements at all these different levels of English proficiency, decreases by a tenth of unit for each year past age-at-arrival nine (Row A4). This is a sizable effect—about a sixth of a standard deviation in the English ordinal measure among immigrants from non-English-speaking countries for each year past age-at-arrival nine. Someone arriving at age 14 would have an English ordinal measure that is lower by half a unit.

III. Age at Arrival, Anglophone Origin, and Social Outcomes

Compared to immigrants with English-speaking countries of origin, immigrants from non-English-speaking countries show substantial age-at-arrival effects for the social outcomes we consider below. This is evident in Figure 2, where, for several of these outcomes, we graph the mean by age at arrival and English/non-English origin.[13] Panel A shows whether the immigrant is currently married with his/her spouse present. Earlier arrivers show essentially similar marriage rates across language-origin groups while later arrivers from non-English-speaking countries are more likely to be married. For Panels B through D, we examine spouse's English-speaking ability, number of children, and fraction of one's "neighborhood" from the same country of birth. Again, outcomes look similar across language-origin groups for the earlier ages at arrival and then diverge, with later arrivers from non-English-speaking countries marrying people with lower English proficiency, having more children and living in ethnic

[13] To save space, we do not make graphs for every outcome and instead choose an outcome for each of the four broad categories of outcomes we consider below—marital status, spousal characteristics, fertility and residential location. However, we do present reduced-form coefficients, which are more succinct summaries of the information in such graphs, for every outcome below in Table 2.

enclaves.[14]

We can summarize the relationship between age at arrival and the difference in social outcomes by language-origin group in a regression framework. To do so, we estimate Equation 2 where the dependent variable is a social outcome rather than English proficiency; this is a reduced-form equation. We do this for all the social outcomes we consider below, not only the four in Figure 2. The results are displayed in Panels B-G of Table 2. As an example, for each year past age nine that an immigrant from a non-English-speaking country arrives, he/she is 1.12 percentage points more likely to be married with spouse present and has 0.046 fewer children.

We attribute these differential age-at-arrival effects to language proficiency. First, recall the coincidence of the English-language effect with the critical period of language acquisition (Figure 1). Second, note the similarity in shape between the curve for English proficiency and the curves for social outcomes in Figure 2.[15] These empirical observations suggest the following causal mechanism: childhood immigrants with first exposure to English after the critical period attain poorer English proficiency as adults, which in turn influences their social outcomes.

The reduced-form effects reported in Table 2 reveal some significant effects of English proficiency on marriage and fertility outcomes. Since English proficiency is decreasing in age at arrival, a negative reduced-form effect is interpreted as a positive effect of English proficiency and a positive reduced-form effect is interpreted as a negative effect.

IV. Instrumental-Variables Estimates

It may be more appealing to rescale these reduced-form estimates to have a direct

[14] It is interesting to note that the age-at-arrival profile for these social outcomes for immigrants from English-speaking countries is not always flat. That is, for some outcomes, there are differences between younger and older arrivers among immigrants from English-speaking countries. These are what we call the non-language-related age-at-arrival effects, and it is the possibility of such effects that prompts us to use age at arrival interacted with being born in a non-English-speaking country as the exclusion restriction rather than age at arrival itself.

[15] To save space, we do not provide graphs of the difference in means for the outcomes in Figure 2. These graphs have the same basic shape as the profiles for immigrants from English-speaking countries alone because the profiles for immigrants from English-speaking countries are always flatter.

interpretation as an effect of English proficiency, i.e., instead of answering what the change in outcome associated with coming one year past age at arrival nine, we answer what is the effect of a unit-change in the English ordinal measure. This is why we proceed to instrumental-variables estimation. Consider the following regression model:

$$y_{ija} = \alpha + \beta\, ENG_{ija} + \delta_a + \gamma_j + \mathbf{w}_{ija}'\rho + \varepsilon_{ija} \qquad (3)$$

for individual i born in country j arriving in the U.S. at age a. y_{ija} is the outcome, ENG_{ija} is a measure of English-language skills (the endogenous regressor), δ_a is a set of age-at-arrival dummies, γ_j is a set of country-of-birth dummies and $\boldsymbol{w}_{ija}$ is a vector of exogenous explanatory variables (e.g., age and sex). Because English skills are endogenous, we cannot obtain unbiased estimates of Equation 3 using OLS. Instead, we use k_{ija}, the excess age-at-arrival effect for non-English-origin immigrants, as an instrumental variable to identify the effect of English-language skill (the β parameter). Since Equation 3 is just-identified, then the indirect least squares estimate (which is the reduced-form coefficient divided by the first-stage coefficient) is identical to the two-stage least squares (2SLS) estimate. We will be using the ordinal measure of English proficiency as the endogenous regressor, and so the coefficient for the first stage equation is reported in Table 2, Row A4. This coefficient is -0.1043, so the 2SLS estimates will be opposite in sign and almost ten times the magnitude compared to the reduced-form effects shown in Table 2, Panels B-G.

In Table 3, we display the OLS and 2SLS estimates of the effect of English proficiency on marital status. Using a sample containing both men and women, 2SLS estimates suggest that English proficiency significantly decreases the probability of being currently married (Column 2, Row A1). This is largely attributable to English proficiency significantly increasing the probability of divorce for women (Row A2) and decreasing the probability of ever marrying (up to April 2000, when the 2000 Census was taken) for men (Row A3).[16] We do not have completed marriage histories for people in our sample—we only observe their marital status at a single point in time—so it is possible that men are only postponing marriage rather than foregoing it altogether. However, it is clear that English proficiency raises the likelihood of ever divorcing. Perhaps English proficiency improves outside opportunities to such an extent that immigrants exit marriages at a lower threshold of marital discord. Alternatively, it could be that greater English proficiency engenders higher expectations of one's own spouse or greater acceptance of the American society's relatively liberal attitude toward divorce.[17]

In Panels B-D, we consider how English proficiency affects the spousal characteristics. This analysis is restricted to those childhood immigrants who are married with spouse present.[18] Panel B shows the effect of English proficiency on the nativity and ethnicity of the spouse. When one's own English proficiency is higher, one's spouse's English proficiency is higher as well (Row B1).[19] Indeed, for men, the coefficient is approximately one, suggesting perfect assortative matching on English-language skills. The effect for women is significantly lower, though still strong: a one-unit increase in own English proficiency leads a woman to choose a spouse with 0.6 units higher English proficiency. More English-proficient people are

[16] The difference in effect by gender is significant at the 5% level for the outcome ever married, but not for the outcome being divorced.

[17] The 2002 National Survey of Latinos provides evidence that attitude toward divorce is associated with English proficiency (Robert Suro et al., 2002). Respondents were asked whether they found divorce to be "acceptable" or "unacceptable" according to their own values and morals, and more-English-proficient people were more likely to answer acceptable. Specifically, 47% of Spanish-dominant Hispanics, 63% of bilingual Hispanics and 67% of English-dominant Hispanics thought divorce was acceptable; the rate was 72% among non-Hispanics.

[18] In the census data, only if the husband and wife live in the same household can spousal characteristics be measured. This has the potential to cause sample-selection bias because English proficiency significantly increases divorce rates, but we lack data on the characteristics of the former spouses of divorced individuals. We argue that this potential bias is unlikely to materially change our findings. First, the magnitude of the effect of English proficiency on divorce is not sufficiently large to overturn the result below that intermarriage increases, and spousal schooling and labor market outcomes improve, with English skill. Second, given that divorce rates are higher among intermarried couples, and that U.S. natives (the most common nationality of spouses of intermarried immigrants) are more educated and earn more than immigrants, if we had included data on the spousal characteristics of divorced individuals, then we might expect even larger magnitudes for the effect of English on spousal characteristics than we have estimated here.

[19] Because this particular regression includes English proficiency on both the left- and right-hand sides, the instrumental-variables coefficient is presumably inflated somewhat by Manski's "reflection" effect.

significantly more likely to marry a U.S. native (Row B2), and this comes at the direct expense of marrying someone born in the same country (Row B3). They are somewhat less likely to marry someone of the same ancestry as well (Row B4), although the smaller magnitude in Row B4 compared to Rows B2 and B3 suggests that some of the U.S. natives they are marrying share their ancestry.[20] For example, English-proficient Mexican immigrants are more likely to marry U.S. natives, some of whom may be of Mexican heritage.

Better English skills lead immigrants to have younger spouses, and this effect comes primarily from the women (Row C1). That is, when a woman is more English proficient, she chooses a younger husband compared to a woman who is less English proficient. This is consistent with the idea that more traditional marriages have a larger age gap between husband and wife (with the husband being several years older), and English proficiency reduces this age gap. (Note these regressions already contain full sets of age and age-at-arrival dummies, so these results are not mechanical.)

More English-proficient people have spouses who are more educated (Row C2). A one-unit increase in English skill raises spousal education by 2.4 years. For the purpose of comparison, in this sample of married couples, a one-unit increase in the English ordinal measure raises own education by 3.5 years. That the effect of English-language skills on one's own education is so similar to the effect on one's spouse's education is indicative of strong assortative matching. Additional analysis reveals that much of the increase in spousal education derives from higher likelihood of finishing high school and attending some college, which parallels estimates for own schooling.

In Panel D, we estimate that better English leads individuals to choose spouses who have

[20] The Census asks "What is this person's ancestry or ethnic origin" and allows for two responses. Row B4 reports the results when we use only the first response, i.e., the outcome measure takes on the value one when the respondent and the spouse both have the same primary ancestry. Results are similar when we define the couple being of the same ancestry as either of their responses matching.

better labor market outcomes. More English-proficient people have spouses who earn 30% more conditional on working (Row D1); in this sample of married couples, a one-unit increase in the English ordinal measure raises own wages by 45%, with the higher estimate not surprising due to the larger effect on own education compared to spousal education. Their spouses are also more likely to work (Row D2); this is similar to the effect on own propensity to work. The estimated effects on working are driven by the wives; husbands' participation in the labor force is high and not sensitive to language skills. Given that husbands' participation in the labor force is high and that wives' participation is increasing in English proficiency, English-proficient people are much more likely to be in marriages in which both the husband and the wife work (Row D3).

We examine fertility outcomes in Table 4. The 2000 Census enables us to construct fertility measures based on the number of children residing in the same household.[21] As above, we consider both men and women; however, because children are more likely to be in the same household with their mothers than their fathers, the results for women are more straightforward to interpret as effects on fertility. We find that English proficiency reduces fertility, especially along the intensive margin. In Row 1, we estimate the impact of English proficiency on the total number of children in the household, and find a 2SLS effect of -0.4. This reduction in number of children is partly due to more English-proficient people having no children at all, especially the men (Row 2). However, looking at these two fertility outcomes among married couples reveals that the previous effect for men at the extensive margin is due to English proficiency making men more likely to remain single (Rows 3 and 4). Indeed, among married couples, English

[21] Unfortunately this means that children who have left the household will not be counted. This will bias our results if the age distribution of children and probability of child leaving parental household conditional on age depend on parental English proficiency. To guard against this possibility, we replicate our design (results available upon request) using the 1990 Census, which offers a better measure of fertility—children ever born to a woman. Analysis of both fertility measures (children ever born and children residing in the same household) yields the same conclusions about the impact of English proficiency on fertility. Moreover, results using the resident-children measure agree across the two different censuses. This raises our confidence that the fertility results using 2000 Census data truly relate to fertility and are not seriously biased by children endogenously leaving the parental household.

proficiency affects only the intensive margin of fertility—a significant reduction on the number of children is paired with no effect on whether a couple has at least one child. Given that we do not observe completed fertility histories for people in our sample, it is fair to ask whether couples are really reducing their fertility or just postponing it. On average though, delaying childbirth will lead to lower completed fertility; among the reasons is that fecundity decreases with woman's age (Philip S. Morgan, 1996).

Rows 5 and 6 look at measures of single parenthood, which is of concern to some policymakers. We do not find any evidence that English proficiency raises single parenthood; there is no evidence that English proficiency raises out-of-wedlock births or that the significant increase in divorce caused by English proficiency creates more single-parent households.

In Table 5, we examine residential location outcomes. In particular, we wish to measure the extent to which an immigrant lives in an ethnic enclave. The public-use 2000 Census data that we use are not ideal for studying residential choice decisions because they do not provide detailed information about the neighborhood that a person resides to preserve respondent privacy. The lowest level of geographic aggregation that we can measure is something called a PUMA (public-use microdata area), which is an area containing at least 100,000 people. A more accurate characterization of one's neighborhood would contain fewer people, but, given the data limitations, we form a couple of variables intended to capture the idea of an ethnic enclave. One set of measures that we use is based on the fraction of the population in one's PUMA that shares the same country of birth as the childhood immigrant (Panel A).[22] Of course there are people of the same background who are born in the U.S. (e.g., U.S.-born Mexicans may have many

[22] Specifically, we use the share of people aged 20 to 60 in the 5% PUMS file whose birthplace matches respondent's birthplace. Recognizing that an enclave is really where there is a big concentration of people from own group, we also tried nonlinear functions of the fraction. Here, we show results not only for when the fraction itself is the outcome, but also for when a dummy for whether an individual lives in a PUMA that has a fraction from same country of birth that is above the mean fraction for all immigrants from that same country of birth (this latter measure picks up whether for someone from your country, you tend to live in a neighborhood with an above-average number of fellow countrymen).

similarities to Mexicans born in Mexico), so a second set of measures used is based on the fraction of one's PUMA that shares the same primary ancestry (Panel B).[23] A larger fraction from the same country or ancestry may be associated with being in a larger ethnic community and greater likelihood of living in an ethnic enclave. For both sets of measures, we find negative effects though the 2SLS effects are typically not significant. Thus, there is only weak evidence of English proficiency affecting enclave residence for the sample overall, though the effect appears to be heterogeneous—English proficiency significantly decreases enclave residence for women (Table 5, Column 4)[24] and non-Mexican immigrants from non-English-speaking countries (Table 6, Column 5, Panel F which we discuss below).

Ex ante, we could not have known whether OLS would be upward-biased or downward-biased. This motivated us to use an instrumental-variables strategy to obtain a consistent estimate of the effect of English proficiency. Our general finding is that the OLS and 2SLS estimates have the same sign, with the OLS estimates tending to be smaller in magnitude.[25] At first glance, this would seem at odds with a story of endogeneity bias in which higher ability immigrants both learn more English and obtain better outcomes in the labor and marriage markets, for example. However, Christian Dustmann and Arthur van Soest (2002) argue that the categorical measure of language proficiency employed by various surveys including the U.S. Census is characterized by substantial measurement error. It is well known that 2SLS can correct for measurement error as well as endogeneity bias. Accordingly, using an alternative measure of English proficiency for validation, Bleakley and Chin (2004) find that the downward

[23] Panel B's enclave measures are based on the primary ancestry reported by respondents. Results are similar when we use measures that incorporate the secondary ancestry as well, e.g., share of respondent's PUMA's population whose first or second ancestry matches respondent's first or second ancestry.
[24] The difference in effect by gender is significant at the 5% level for all enclave residence measures except the one in Row A2.
[25] According to Hausman tests, the 2SLS estimate and the OLS estimate differ at the 5% level of significance for the following outcomes: all three marital status measures; spouse is U.S.-born; spouse has same country of birth; spouse years of schooling; and number of children living in the household among all individuals and among married couples.

bias caused by classical measurement error outweighs the upward bias due to an "ability bias"-type story.

V. Robustness Checks and Extensions

In this section, we consider several alternative hypotheses for the results from above on English-speaking ability and social outcomes. For the 2SLS estimate, we interpret the age-at-arrival effect for immigrants from non-English-speaking countries that is in excess of the age-at-arrival effect for immigrants from English-speaking countries as the causal effect of English proficiency. However, if non-language age-at-arrival effects differ between the two groups of immigrants, then our strategy to identify the effect of English proficiency is invalid. For example, English-speaking countries and non-English-speaking countries may differ in ways that affect the assimilation process of immigrants in the U.S. To assess this potential problem, we perform a variety of specification checks.

First, it is possible that immigrants from non-English-speaking countries exhibit a stronger age-at-arrival effect simply because immigrants from poorer countries face additional barriers to adaptation and that these barriers increase in severity as a function of age at arrival. This is plausible because non-English-speaking countries tend to be poorer than English-speaking countries. Richer countries might have better school systems. If there are different returns associated with the schooling obtained in a non-English-speaking country versus an English-speaking one, the 2SLS estimate using the interaction as the identifying instrument may reflect not only differential English-language skills but also differential returns to origin-country schooling. To address this, we incorporate data on per capita GDP in 1980 from the Penn World Tables (Robert Summers and Alan Heston, 1988). We include as a control variable an interaction between age at arrival and per capita GDP in the country of birth.[26] The estimation results, shown in Column 2 of Table 6, are similar to the base results (copied in Column 1 from Tables 3-5).

Second, the age-at-arrival effect could depend on the fertility rate in the origin country. Assimilation to U.S. norms would mean a reduction in fertility for people from higher-fertility countries but an increase in fertility for people from lower-fertility countries. The fertility rate in the U.S. is higher than in most other industrialized countries, but lower than in most developing countries, and English-speaking countries are more likely to be industrialized. Thus, immigrants from English-speaking countries may not properly control for the non-language age-at-arrival effects on fertility experienced by immigrants from non-English-speaking countries. To address this potential source of bias, we incorporate data on total fertility rate in 1982 from the *World Development Indicators* CD-ROM (World Bank, 2005). We include as a control variable an interaction between age at arrival and total fertility rate in the country of birth. The estimation results, shown in Column 3, are similar to the base results.

Finally, English-speaking countries might have greater cultural and institutional similarity to the U.S., making adjustment easy for immigrants from these countries irrespective of age at arrival. In contrast, immigrants from non-English-speaking countries encounter both a foreign language and foreign culture, so even ignoring the language, there is more to adjust to for the older arrivers. To address this concern, we restrict analysis to groups of countries that might be more similar to each other. In Column 4, we drop immigrants from Canada.[27] They account for almost one third of the observations of immigrants from English-speaking countries, yet they may be poor controls for the assimilation process of the average immigrant due to Canada's

[26] Results controlling for an interaction between age at arrival and school quality in the country of birth are similar, and are not displayed here. We used two measures of school quality in turn: school expenditures per child and teacher-pupil ratio at the primary level from Jong-Wha Lee and Robert Barro (1997).

[27] Results dropping immigrants from the United Kingdom, Australia and New Zealand in addition to those from Canada are similar, and are not displayed here.

geographic proximity to the U.S. and status as a former British colony. The results are broadly similar to those in Column 1.

In Column 5, we drop immigrants from Mexico. They account for 29% of the observations of immigrants from non-English-speaking countries. By dropping them, we can explore whether the estimated effect of English proficiency is driven by Mexicans alone, or whether the effect is common to other groups as well. Although most of the results are qualitatively similar to the base results, several differences should be noted. Now a one-unit increase in English proficiency generates a larger increase in the probability of marrying a U.S. native and a larger decrease in the probability of marrying a fellow countryman. Also, the negative effect on the probability of having a spouse with the same ancestry is now significant and larger in magnitude (the point estimate is -0.51 compared to the base result of -0.18). Moreover, the effect on living in an ethnic enclave is now always significant and larger in magnitude. We must recognize that the estimates are imprecise, but the following story seems plausible. Mexicans are much more numerous than other immigrant or ancestry groups in the U.S., and due to their numbers and relative geographic concentration have more and deeper ethnic communities. This means that a Mexican immigrant who is English proficient has a larger chance of finding a mate satisfying the education and earnings requirements who is also of Mexican ancestry. Additionally, because the membership is so large and varied, Mexican enclaves may be attractive even for English-proficient Mexicans to live in. On the other hand, non-Mexican immigrants typically have to marry someone born in a different country or of a different ancestry to satisfy their requirements. Their group's enclaves may be small and only be attractive for newly arrived immigrants. A different story that is also consistent with these results is that Mexicans have a stronger preference to marry other Mexicans and live with each other regardless of English proficiency.

Overall, Table 6 suggests that our main findings are robust to changes to sample or specification that might make the immigrants from English-speaking countries better controls for the non-language age-at-arrival effects experienced by immigrants from non-English-speaking countries.

As an extension to our analysis, we consider whether the estimated effects of English proficiency on marriage, fertility and ethnic enclave residence are working through differences in education. For example, is the reduction in fertility fully explained by the rise in the opportunity cost of female time associated with an increase in English proficiency, or is there evidence of other mechanisms at work? For suggestive evidence on this point, we estimated the same specifications as before but add years of schooling as a regressor.[28] We find that the effects of English proficiency on marital status, fertility, and spousal origin do not change much after controlling for education. This suggests that the additional education attained as a result of better English is not the central channel for these results, leaving room for some other channels for the effect of English proficiency, such as enabling communication (thus increasing the pool of suitors), social learning (discovering and adopting U.S. cultural norms), and raising female bargaining power (through improving exit options for women disproportionately). In contrast, for spousal education and wages, the coefficients for English proficiency decline markedly after controlling for education.

VI. Conclusion

Among immigrants who came to the U.S. during childhood, we observe that age at arrival is also related to various social outcomes. These relationships are especially marked for those immigrants from non-English-speaking countries, for whom arriving later predicts 10%

[28] These results are available upon request. This evidence is only suggestive because education, like English proficiency, is likely endogenous.

higher marriage rate and 0.5 more children, over and above late-arriving immigrants from Anglophone countries. These facts complement earlier findings (Bleakley and Chin, 2004) that immigrating earlier, particularly from non-Anglophone countries, was associated with higher adult incomes and more years of schooling (by up to 15% and 2 years, respectively).

Guided by language-acquisition theory, we developed an instrumental variables strategy to identify the causal effect of U.S. immigrants' English proficiency on various social outcomes. We find that English proficiency raises the probabilities of being divorced, marrying a U.S. native, having a more educated and higher-earning spouse, having fewer children, and, for some groups, living outside of ethnic enclaves. These results indicate that English-language skills have an important role in the process of social assimilation, and furthermore that the marriage market for immigrants is characterized by strongly assortative matching by English proficiency, education and labor market outcomes.

These results help us understand the household environment in which the children of immigrants grow up. Immigrants with higher English proficiency have spouses who are U.S. natives, more educated and earn more. This means that marriage decisions magnify existing differences across individuals along linguistic lines. For example, when someone marries a U.S. native, his/her use and knowledge of English will grow. Also, when someone marries another higher earner, total family income will rise. Bleakley and Chin (2008) find that the English proficiency of immigrant parents has a significant benefit for English proficiency and educational outcomes of their U.S.-born children. Likely, an important mediator for these intergenerational effects is the family structure. First, children with one parent with low English proficiency are more likely to have the other parent be less English-proficient too, which means lower education and earnings in the family on average. Second, these children have more siblings on average, which can have consequences for children's outcomes.

These results call into question the commonplace hypothesis that immigrants' extent of English proficiency and social assimilation is decided by their culture or preferences. Constraints have an important role in immigrant assimilation too. Even immigrants with strong preferences to integrate more fully into U.S. society may be constrained from doing so by their human capital; the example from this paper is that they may have arrived past the critical period so their English-language skills are worse, leading them to act less American than they wish.

We close with a few comments about policy. First, the constraint on language learning that we consider above has a high (shadow) price associated with it. Nevertheless, it would be difficult for immigrants to work around this constraint directly inasmuch as the timing of migration is largely determined by administrative rationing under current U.S. policy. A system that favors migrating when one's children are younger (perhaps in the way the current Canadian "points" system favors or disfavors certain demographic characteristics) might improve the welfare of would-be immigrants, although it would have to be artfully designed so as not to encourage gaming the system. Second, we do not propose to manipulate language policy in order to attain certain marriage or fertility outcomes. However, language policy is often manipulated for the sake of improving education and earnings outcomes, and this study points out that there will be concomitant effects on marriage and fertility. Third, our results suggest further examination of subsidized language-instruction programs, although such programs are probably poor substitutes for immigrating earlier in childhood, given the importance of the critical-period constraint, even in spite of language instruction already being supplied (probably elastically) by the market.

References

Birdsong, David. 2006. "Age and second language acquisition and processing: A selective overview," *Language Learning* 56:9-49.

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

Bleakley, Hoyt and Aimee Chin. 2008. "What Holds Back the Second Generation? The Intergenerational Transmission of Language Human Capital among Immigrants," *Journal of Human Resources* 43(Spring):267-298.

Capps, Randy, Michael Fix and Jane Reardon-Anderson. 2003. "Children of Immigrants Show Slight Reductions in Poverty, Hardship," In *Snapshots of American Families III*, No. 13, Washington, D.C.: The Urban Institute.

Dávila, Alberto and Marie T. Mora. 2001. "The Marital Status of Recent Immigrants in the United States in 1980 and 1990," *International Migration Review* 35(2):506-524.

Duncan, Brian and Stephen Trejo. 2007. "Ethnic Identification, Intermarriage and Unmeasured Progress by Mexican Americans." In *Mexican Immigration to the United States*, ed. George Borjas, 229-267. Chicago: National Bureau of Economic Research and the University of Chicago Press.

Dustmann, Christian and Arthur van Soest. 2002. "Language and the Earnings of Immigrants," *Industrial and Labor Relations Review* 55(3):473-492.

Fernández, Raquel and Alessandra Fogli. 2006. "Fertility: The Role of Culture and Family Experience," *Journal of the European Economic Association*, 4(2-3):552-561.

Funkhouser, Edward and Fernando A. Ramos. 1993 "The Choice of Migration Destination: Dominican and Cuban Immigrants to the Mainland United States and Puerto Rico," *International Migration Review* 27(3):537-556.

Kominski, Robert. 1989. "How Good Is 'How Well'? An Examination of the Census English-Speaking Ability Question." Paper presented at the 1989 annual meeting of the American Statistical Association, Washington..

Lazear, Edward P. 2007. "Ethnic Identification, Intermarriage and Unmeasured Progress by Mexican Americans." In *Mexican Immigration to the United States*, ed. George Borjas, 107-122. Chicago: National Bureau of Economic Research and the University of Chicago Press.

Lee, Jong-Wha and Robert J. Barro. 2001. "Schooling Quality In A Cross-Section Of Countries," *Economica*, 68(271):465-488. (data from http://www.nber.org).

Lenneberg, Eric H. 1967. *Biological Foundation of Language*, New York: Wiley & Sons.

Meng, Xin and Robert G. Gregory. 2005. "Intermarriage and the Economic Assimilation of Immigrants," *Journal of Labor Economics,* 23(1):135-175.

Morgan, S. Philip. 1996. "Characteristic Features of Modern American Fertility," *Population and Development Review,* 22(Supplement):19-63.

Newport, Elissa L. 2002. "Critical Periods in Language Development." In *Encyclopedia of Cognitive Science*, ed. L. Nadel. London: Macmillan Publishers Ltd./ Nature Publishing Group.

Ruggles, Steven, Matthew Sobek, Trent Alexander, Catherine A. Fitch, Ronald Goeken, Patricia Kelly Hall, Miriam King, and Chad Ronnander. 2004. *Integrated Public Use Microdata Series: Version 3.0* [Machine-readable database], Minneapolis, MN: Minnesota Population Center [producer and distributor]. (http://www.ipums.org).

Sorenson, Ann Marie. 1988. "The Fertility and Language Characteristics of Mexican-American and Non-Hispanic Husbands and Wives," *Sociological Quarterly,* 29(1):111-130.

Stevens, Gillian and Gray Swicegood. 1987. "The Linguistic Context of Ethnic Endogamy," *American Sociological Review,* 52(1):73-82.

Summers, Robert and Alan Heston. 1988. "A New Set of International Comparisons of Real Product and Price Levels Estimates for 130 Countries, 1950-1985," *Review of Income and Wealth,* 34(1):1-25. (Penn World Tables (Mark 5.6a) data from http://www.nber.org).

Suro, Robert, Mollyann Brodie, Annie Steffenson, Jaime Valdez and Rebecca Levin. 2002. "Pew Hispanic Center/Kaiser Family Foundation 2002 National Survey of Latinos." Pew Hispanic Center Report, December 2002.

Swicegood, Gray, Frank D. Bean, Elizabeth Hervey Stephen and Wolfgang Opitz. 1988. "Language Usage and Fertility in the Mexican-Origin Population of the United States," *Demography,* 25(1):17-33.

Toussaint-Comeau, Maude and Sherrie L.W. Rhine. 2004. "Tenure Choice with Location Selection: The Case of Hispanic Neighborhoods in Chicago," *Contemporary Economic Policy* 22(1):95-110.

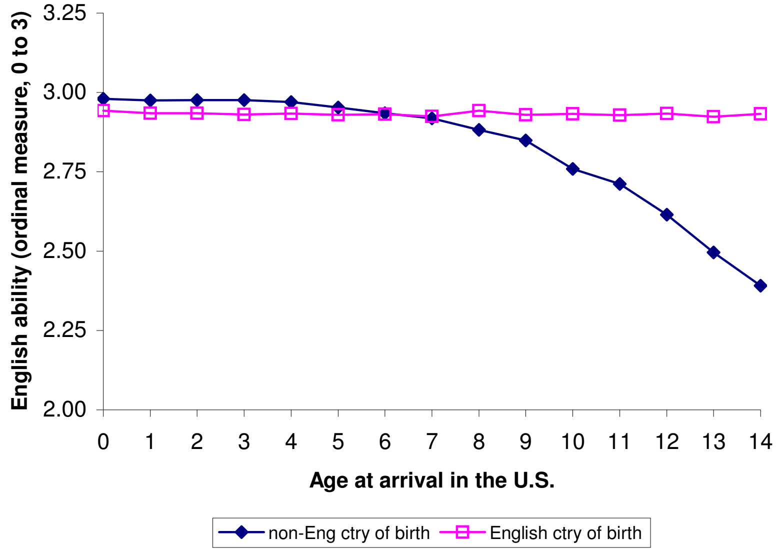
Weil, David N. 2005. *Economic Growth*, U.S.: Pearson Addison Wesley.

World Almanac and Book of Facts, 1999, New York: World Almanac Books.

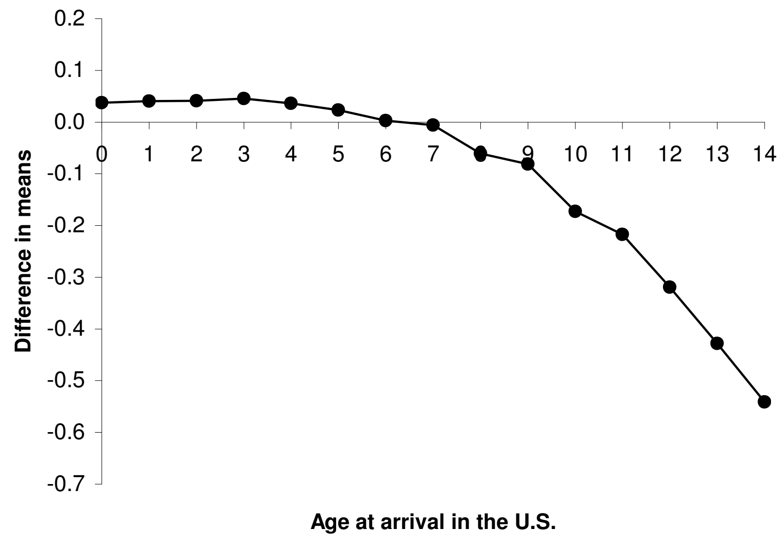
World Bank, *World Development Indicators 2005* [CD-ROM]. 2005. Washington, DC: World Bank [Producer and Distributor].

Figure 1. English-Speaking Ability by Age at Arrival

Panel A. Regression-Adjusted Means

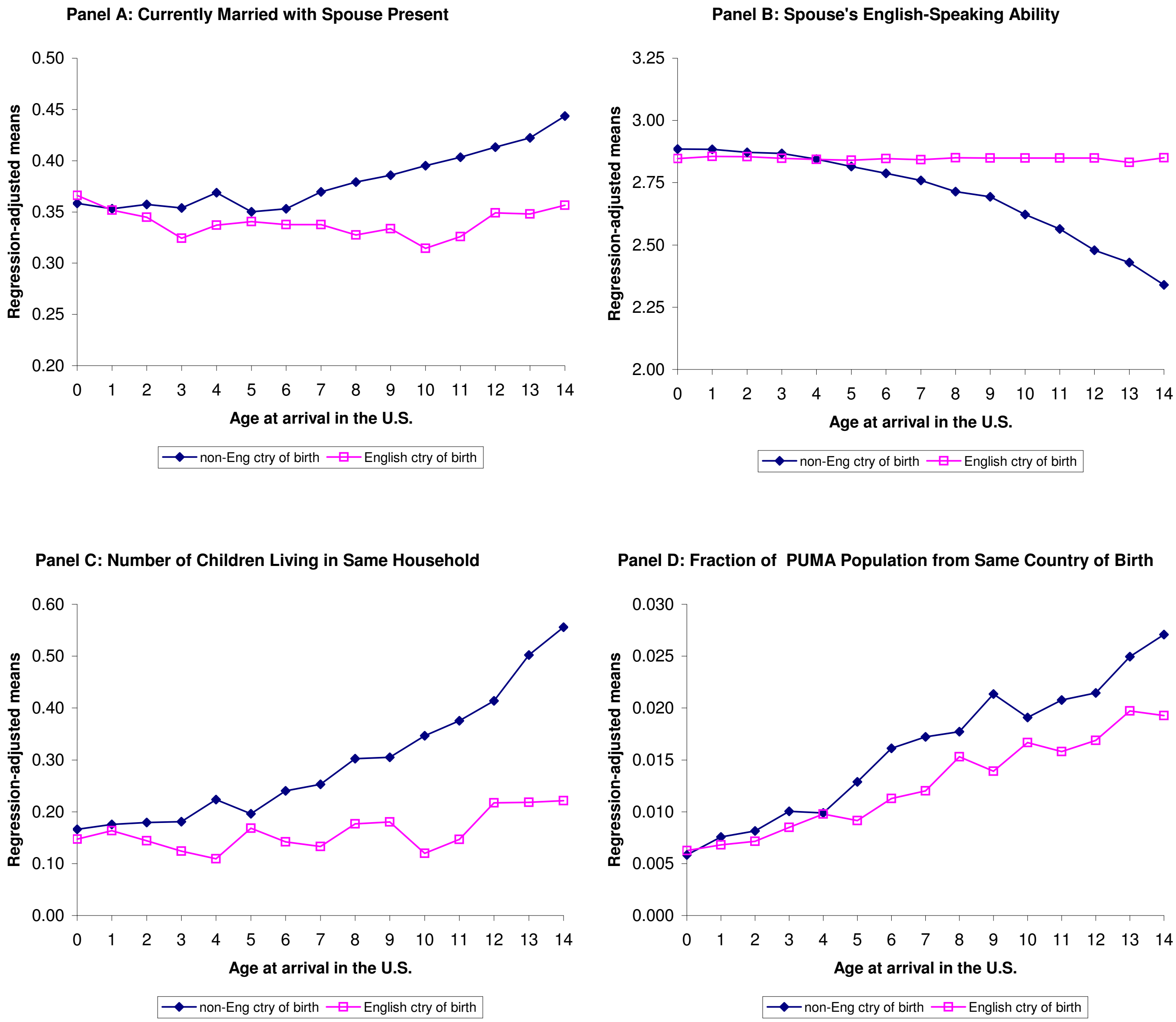


Panel B. Difference in Means



Notes: Data are from the 2000 IPUMS. Sample size is 191,534 (composed of people who immigrated to the U.S. before age 15 and are currently aged 25-55, and with nonmissing English variable). In Panel A, displayed for each age at arrival is the mean English-speaking ability. In Panel B, displayed for each age at arrival is the difference in mean English-speaking ability between people from non-English-speaking countries and people from English-speaking countries. Means are weighted by IPUMS weights, and regression-adjusted for age, race, Hispanic and sex dummies. The race categories used were White, Black, Asian & Pacific Islander, Multiracial and Other. The English-speaking ability ordinal measure is defined as: 0 = no English, 1 = not well, 2 = well and 3 = very well.

Figure 2. Select Outcomes by Age at Arrival



Notes: Data are from the 2000 IPUMS. Panels A , C and D use data for all childhood immigrants and Panel B uses data for the subset that is currently married with spouse present. Means are weighted by IPUMS weights, and regression-adjusted for age, race, Hispanic and sex dummies.

Table 1. Descriptive Statistics

	Born in non-English-speaking country			Born in English-speaking country		
	total	arrived aged 0-9	arrived aged 10-14	total	arrived aged 0-9	arrived aged 10-14
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Regressors						
English-speaking ability ordinal measure	2.719	2.872	2.441	2.980	2.981	2.979
	(0.619)	(0.420)	(0.797)	(0.167)	(0.167)	(0.170)
Age	36.549	36.839	36.025	38.403	38.906	37.014
	(8.256)	(8.357)	(8.044)	(8.367)	(8.387)	(8.153)
Female	0.500	0.512	0.478	0.528	0.513	0.569
	(0.500)	(0.500)	(0.500)	(0.499)	(0.500)	(0.495)
White	0.554	0.609	0.454	0.666	0.766	0.390
	(0.497)	(0.488)	(0.498)	(0.472)	(0.424)	(0.488)
Black	0.030	0.031	0.030	0.247	0.164	0.478
	(0.171)	(0.172)	(0.170)	(0.432)	(0.370)	(0.500)
Asian/Pacific Islander	0.123	0.106	0.154	0.028	0.024	0.038
	(0.328)	(0.307)	(0.361)	(0.164)	(0.153)	(0.192)
Other single race	0.240	0.201	0.310	0.019	0.016	0.028
	(0.427)	(0.401)	(0.462)	(0.136)	(0.124)	(0.164)
Multiracial	0.053	0.053	0.053	0.040	0.031	0.066
	(0.224)	(0.224)	(0.223)	(0.196)	(0.173)	(0.248)
Hispanic	0.520	0.452	0.644	0.012	0.012	0.011
	(0.500)	(0.498)	(0.479)	(0.109)	(0.110)	(0.106)
Years of schooling	13.138	13.753	12.014	14.527	14.593	14.342
	(3.461)	(2.956)	(3.994)	(2.452)	(2.446)	(2.459)
Panel B: Marriage Outcomes						
Is currently married with spouse present	0.604	0.588	0.632	0.561	0.584	0.497
	(0.489)	(0.492)	(0.482)	(0.496)	(0.493)	(0.500)
Is currently divorced	0.097	0.107	0.077	0.120	0.123	0.112
	(0.296)	(0.309)	(0.267)	(0.325)	(0.328)	(0.315)
Has ever married	0.767	0.757	0.787	0.736	0.755	0.680
	(0.422)	(0.429)	(0.409)	(0.441)	(0.430)	(0.466)
Spouse English-speaking ability ordinal measure	2.588	2.758	2.301	2.979	2.981	2.973
	(0.765)	(0.599)	(0.916)	(0.170)	(0.164)	(0.189)
Spouse is US-born	0.494	0.632	0.261	0.804	0.859	0.622
	(0.500)	(0.482)	(0.439)	(0.397)	(0.348)	(0.485)
Spouse has the same country of birth	0.393	0.265	0.609	0.094	0.055	0.223
	(0.488)	(0.442)	(0.488)	(0.292)	(0.228)	(0.416)
Spouse has the same primary ancestry	0.543	0.459	0.681	0.245	0.218	0.331
	(0.498)	(0.498)	(0.466)	(0.430)	(0.413)	(0.471)
Spouse age	38.077	38.479	37.399	40.491	40.823	39.410
	(9.069)	(9.069)	(9.029)	(8.849)	(8.806)	(8.904)
Spouse years of schooling	13.016	13.645	11.936	14.578	14.647	14.351
	(3.704)	(3.267)	(4.135)	(2.530)	(2.525)	(2.534)
Spouse log(wages last year)	10.201	10.265	10.078	10.370	10.371	10.368
	(0.991)	(0.988)	(0.984)	(1.025)	(1.042)	(0.965)
Spouse worked last year	0.825	0.848	0.786	0.883	0.884	0.879
	(0.380)	(0.359)	(0.410)	(0.322)	(0.321)	(0.326)
Both worked last year	0.701	0.733	0.646	0.781	0.784	0.771
	(0.458)	(0.442)	(0.478)	(0.414)	(0.412)	(0.421)

Notes: The table continues on the next page.

Table 1. Descriptive Statistics (Continued)

	Born in non-English-speaking country			Born in English-speaking country		
	total (1)	arrived aged 0-9 (2)	arrived aged 10-14 (3)	total (4)	arrived aged 0-9 (5)	arrived aged 10-14 (6)
Panel C: Fertility Outcomes						
Number of children living in same household	1.246	1.141	1.436	0.974	0.969	0.986
	(1.346)	(1.279)	(1.440)	(1.178)	(1.175)	(1.187)
Has a child living in same household	0.580	0.554	0.628	0.506	0.502	0.518
	(0.493)	(0.497)	(0.483)	(0.500)	(0.500)	(0.500)
Number of children living in same household, only individuals married with spouse present	1.749	1.626	1.956	1.421	1.400	1.487
	(1.312)	(1.261)	(1.369)	(1.208)	(1.208)	(1.203)
Has a child living in same household, only individuals married with spouse present	0.797	0.772	0.838	0.715	0.703	0.755
	(0.402)	(0.419)	(0.368)	(0.451)	(0.457)	(0.430)
Is a single parent	0.101	0.101	0.101	0.107	0.093	0.144
	(0.301)	(0.301)	(0.301)	(0.309)	(0.291)	(0.351)
Is a never-married single parent	0.029	0.028	0.033	0.038	0.028	0.064
	(0.169)	(0.164)	(0.178)	(0.190)	(0.165)	(0.244)
Panel D: Residential Location Outcomes						
Fraction of PUMA population from same country of birth	0.062	0.051	0.081	0.008	0.006	0.014
	(0.103)	(0.096)	(0.112)	(0.019)	(0.015)	(0.028)
Fraction from same country of birth is above national mean for the country of birth	0.335	0.312	0.377	0.351	0.347	0.359
	(0.472)	(0.463)	(0.485)	(0.477)	(0.476)	(0.480)
Fraction of PUMA population with same primary ancestry	0.114	0.106	0.129	0.044	0.047	0.036
	(0.154)	(0.146)	(0.166)	(0.062)	(0.063)	(0.058)
Fraction with same ancestry primary is above national mean for the primary ancestry	0.364	0.352	0.384	0.362	0.364	0.358
	(0.481)	(0.478)	(0.486)	(0.481)	(0.481)	(0.479)

Notes: The sample consists of individuals from the 2000 1% and 5% PUMS files who are currently aged 25-55, immigrated to the U.S. before age 15 and has nonmissing own age, year of immigration, country of birth and English variables. Total number of observations is 191534 for the English variable, with Columns 1-6 containing 165628, 106890, 58738, 25906, 19217and 6689 observations, respectively. For the spouse-related variables in Panel B and the couple fertility measures in Panel C, statistics are reported for the subsample that is currentllly married with spouse present; total number of observations is 114190 for the spouse age variable, with Columns 1-6 containing 99481, 62794, 36687, 14709, 11323 and 3386 observations, respectively. Panel D uses the PUMA variable, which is only available in the 5% PUMS files, so the number of observations is correspondingly lower. Statistics are weighted by IPUMS weights. The English-speaking ability ordinal measure is defined as: 0 = no English, 1 = not well, 2 = well and 3 = very well.

Table 2. Reduced-form Effects

Dependent variable	Coefficient for identifying instrument	
Panel A: English Proficiency Measures		
1. Speaks English not well or better	-0.0061	**
	(0.0031)	
2. Speaks English well or better	-0.0293	**
	(0.0124)	
3. Speaks English very well	-0.0689	***
	(0.0135)	
4. English-speaking ability ordinal measure	-0.1043	***
	(0.0288)	
Panel B: Marital Status		
1. Is currently married with spouse present	0.0112	***
	(0.0040)	
2. Is currently divorced	-0.0054	***
	(0.0018)	
3. Has ever married	0.0075	***
	(0.0026)	
Panel C: Spouse's Nativity and Ethnicity		
1. Spouse English-speaking ability ordinal measure	-0.0859	***
	(0.0191)	
2. Spouse is US-born	-0.0342	***
	(0.0113)	
3. Spouse has the same country of birth	0.0373	***
	(0.0122)	
4. Spouse has the same ancestry	0.0191	*
	(0.0105)	
Panel D: Spouse's Age and Education		
1. Spouse age	0.0956	***
	(0.0354)	
2. Spouse years of schooling	-0.2493	***
	(0.0721)	

Dependent variable	Coefficient for identifying instrument	
Panel E: Spouse's Labor Market Outcomes		
1. Spouse log(wages last year)	-0.0314	***
	(0.0118)	
2. Spouse worked last year	-0.0082	***
	(0.0031)	
3. Both worked last year	-0.0127	**
	(0.0051)	
Panel F: Fertility		
1. Number of children living in same household	0.0460	***
	(0.0142)	
2. Has a child living in same household	0.0076	**
	(0.0039)	
3. Number of children living in same household, only individuals married with spouse present	0.0435	***
	(0.0139)	
4. Has a child living in same household, only individuals married with spouse present	0.0007	
	(0.0022)	
5. Is a single parent	-0.0022	
	(0.0027)	
6. Is a never-married single parent	0.0003	
	(0.0018)	
Panel G: Residential Location		
1. Fraction of PUMA population from same country of birth	0.0007	
	(0.0007)	
2. Fraction from same country of birth is above national mean for the country of birth	0.0016	
	(0.0068)	
3. Fraction of PUMA population with same primary ancestry	0.0018	
	(0.0013)	
4. Fraction with same ancestry primary is above national mean for the primary ancestry	0.0023	
	(0.0059)	

Notes: The sample is as described in Table 1 notes with modifications as follows: (1) Panels C-E and couple's fertility in Panel F use the subsample that is married with spouse present; and (2) Panel G only uses the 5% PUMS files. The coefficient reported is for the identifying instrument, max(0, age at arrival - 9)×born in a non-English-speaking country. Each coefficient is from a separate OLS regression that is weighted by IPUMS weights and contains dummies for country of birth, age at arrival, age, race (White, Black, Asian, Multiracial and Other), Hispanic origin and sex. Country-of-birth dummies are based on IPUMS detailed birthplace codes. In Panel III, same birthplace means the IPUMS detailed birthplace codes are the same, and same ancestry means that the general codes for first response to ancestry question are the same. Standard errors adjusted for country of birth clusters are shown in parentheses. Asterisks denote significance levels (*=.10, **=.05, ***=.01).

Table 3. Effect of English-Language Skills on Marriage Outcomes

	All Childhood Immigrants		Women		Men	
	OLS (1)	2SLS (2)	OLS (3)	2SLS (4)	OLS (5)	2SLS (6)
Panel A: Marital Status						
1. Is currently married with spouse present	0.008	-0.108 ***	-0.004	-0.076 **	0.019 *	-0.141 ***
	(0.009)	(0.040)	(0.009)	(0.032)	(0.011)	(0.054)
2. Is currently divorced	0.010 ***	0.052 ***	0.015 ***	0.064 **	0.006	0.038 *
	(0.003)	(0.016)	(0.003)	(0.027)	(0.004)	(0.022)
3. Has ever married	-0.002	-0.072 **	-0.001	-0.015	-0.004	-0.134 ***
	(0.007)	(0.028)	(0.005)	(0.028)	(0.009)	(0.050)
Panel B: Spouse's Nativity and Ethnicity						
1. Spouse English-speaking ability ordinal measure	0.512 ***	0.814 ***	0.474 ***	0.618 ***	0.560 ***	1.039 ***
	(0.016)	(0.072)	(0.008)	(0.091)	(0.027)	(0.062)
2. Spouse is US-born	0.105 ***	0.322 **	0.101 ***	0.270 *	0.109 ***	0.386 ***
	(0.009)	(0.129)	(0.011)	(0.142)	(0.006)	(0.121)
3. Spouse has the same country of birth	-0.120 ***	-0.351 ***	-0.118 ***	-0.285 **	-0.122 ***	-0.434 ***
	(0.007)	(0.126)	(0.009)	(0.140)	(0.006)	(0.108)
4. Spouse has the same primary ancestry	-0.076 ***	-0.181	-0.078 ***	-0.171	-0.074 ***	-0.197
	(0.011)	(0.129)	(0.013)	(0.138)	(0.009)	(0.122)
Panel C: Spouse's Age and Education						
1. Spouse age	-0.348 ***	-0.901 **	-0.496 ***	-1.312 **	-0.147 ***	-0.382
	(0.035)	(0.383)	(0.055)	(0.641)	(0.033)	(0.611)
2. Spouse years of schooling	1.353 ***	2.378 ***	1.367 ***	2.214 ***	1.343 ***	2.619 ***
	(0.034)	(0.298)	(0.036)	(0.391)	(0.043)	(0.241)
Panel D: Spouse's Labor Market Outcomes						
1. Spouse log(wages last year)	0.169 ***	0.302 ***	0.166 ***	0.236 **	0.172 ***	0.422 ***
	(0.006)	(0.086)	(0.006)	(0.113)	(0.012)	(0.139)
2. Spouse worked last year	0.039 ***	0.079 ***	0.021 ***	0.022	0.062 ***	0.144 ***
	(0.003)	(0.023)	(0.005)	(0.021)	(0.004)	(0.050)
3. Both worked last year	0.090 ***	0.122 ***	0.106 ***	0.122 ***	0.072 ***	0.118 **
	(0.004)	(0.029)	(0.006)	(0.024)	(0.004)	(0.054)

Notes: Panel A uses the sample described in Table 1 notes. Panels B-D use the subset of the sample that is currently married with spouse present. Same birthplace means the IPUMS detailed birthplace codes are the same, and same ancestry means that the general codes for first response to ancestry question are the same. Each numbered row of each column is from a separate regression that is weighted by IPUMS weights and contains dummies for country of birth, age at arrival, age, race, Hispanic origin and sex. The "2SLS" columns are estimated using 2SLS with max(0, age at arrival - 9)×non-English-speaking country as the identifying instrument. Standard errors adjusted for country of birth clusters are shown in parentheses. Asterisks denote significance levels (*=.10, **=.05, ***=.01).

Table 4. Effect of English-Language Skills on Fertility Outcomes

	All Childhood Immigrants		Women		Men	
	OLS (1)	2SLS (2)	OLS (3)	2SLS (4)	OLS (5)	2SLS (6)
1. Number of children living in same household	-0.106 ***	-0.441 ***	-0.162 ***	-0.472 ***	-0.049 *	-0.403 ***
	(0.021)	(0.075)	(0.019)	(0.078)	(0.026)	(0.102)
2. Has a child living in same household	-0.005	-0.073 *	-0.019 ***	-0.041	0.009	-0.113 **
	(0.004)	(0.038)	(0.003)	(0.035)	(0.007)	(0.049)
3. Number of children living in same household, only individuals married with spouse present	-0.177 ***	-0.410 ***	-0.203 ***	-0.591 ***	-0.148 ***	-0.190
	(0.013)	(0.058)	(0.017)	(0.100)	(0.013)	(0.152)
4. Has a child living in same household, only individuals married with spouse present	-0.019 ***	-0.007	-0.024 ***	-0.020	-0.014 ***	0.008
	(0.004)	(0.020)	(0.004)	(0.021)	(0.004)	(0.035)
5. Is a single parent	-0.001	0.021	-0.001	0.032	-0.001	6.80E-06
	(0.004)	(0.025)	(0.007)	(0.031)	(0.001)	(0.015)
6. Is a never-married single parent	-0.003	-0.002	-0.005	-0.005	-0.002 **	-0.008
	(0.002)	(0.018)	(0.004)	(0.026)	(0.001)	(0.008)

Notes: Rows 1, 2, 5 and 6 use the sample described in Table 1 notes. Rows 3 and 4 use the subsample that is currently married with spouse present. Each numbered row of each column is from a separate regression that is weighted by IPUMS weights and contains dummies for country of birth, age at arrival, age, race, Hispanic origin and sex. The "2SLS" columns are estimated using 2SLS with max(0, age at arrival - 9)×non-English-speaking country as the identifying instrument. Standard errors adjusted for country of birth clusters are shown in parentheses. Asterisks denote significance levels (*=.10, **=.05, ***=.01).

Table 5. Effect of English-Language Skills on Residential Location Outcomes

	All Childhood Immigrants		Women		Men	
	OLS (1)	2SLS (2)	OLS (3)	2SLS (4)	OLS (5)	2SLS (6)
Panel A: Neighborhood Measures based on Fraction of Population in PUMA from Same Country of Birth						
1. Fraction of PUMA population from same country of birth	-0.010 ***	-0.007	-0.012 ***	-0.014 **	-0.008 ***	-0.0002
	(0.001)	(0.008)	(0.002)	(0.006)	(0.001)	(0.011)
2. Fraction from same country of birth is above national mean for the country of birth	-0.050 ***	-0.015	-0.058 ***	-0.038	-0.041 ***	0.007
	(0.007)	(0.068)	(0.006)	(0.054)	(0.009)	(0.089)
Panel B: Neighborhood Measures based on Fraction of Population in PUMA from Same Ancestry						
1. Fraction of PUMA population with same primary ancestry	-0.013 ***	-0.018	-0.017 ***	-0.033 **	-0.009 ***	-0.002
	(0.002)	(0.017)	(0.002)	(0.013)	(0.001)	(0.021)
2. Fraction with same primary ancestry is above national mean for the primary ancestry	-0.041 ***	-0.022	-0.050 ***	-0.088	-0.033 ***	0.056
	(0.007)	(0.062)	(0.007)	(0.062)	(0.008)	(0.072)

Notes: Both panels use the subset of the sample described in Table 1 notes that is from the 5% PUMS files. Same birthplace means the IPUMS detailed birthplace codes are the same, and same ancestry means that the general codes for first response to ancestry question are the same. Each numbered row of each column is from a separate regression that is weighted by IPUMS weights and contains dummies for country of birth, age at arrival, age, race, Hispanic origin and sex. The "2SLS" columns are estimated using 2SLS with max(0, age at arrival - 9)×non-English-speaking country as the identifying instrument. Standard errors adjusted for country of birth clusters are shown in parentheses. Asterisks denote significance levels (*=.10, **=.05, ***=.01).

Table 6. 2SLS Effect of English Using Alternative Samples and Specifications

	Base results (1)	Control for origin GDP × age at arrival (2)	Control for origin fertility × age at arrival (3)	Drop Canada (4)	Drop Mexico (5)
Panel A: Marital Status					
1. Is currently married with spouse present	-0.108 ***	-0.082 **	-0.097 **	-0.109 **	-0.160 **
	(0.040)	(0.032)	(0.048)	(0.048)	(0.073)
2. Is currently divorced	0.052 ***	0.051 ***	0.054 ***	0.049 **	0.071 **
	(0.016)	(0.013)	(0.020)	(0.020)	(0.030)
3. Has ever married	-0.072 **	-0.043 **	-0.057 *	-0.072 **	-0.116 **
	(0.028)	(0.022)	(0.032)	(0.033)	(0.050)
Panel B: Spouse's Nativity and Ethnicity					
1. Spouse English-speaking ability ordinal measure	0.814 ***	0.757 ***	0.824 ***	0.802 ***	1.031 ***
	(0.072)	(0.054)	(0.074)	(0.071)	(0.093)
2. Spouse is US-born	0.322 **	0.261 ***	0.368 **	0.224 **	0.593 ***
	(0.129)	(0.082)	(0.166)	(0.102)	(0.216)
3. Spouse has the same country of birth	-0.351 ***	-0.285 ***	-0.389 **	-0.278 **	-0.591 ***
	(0.126)	(0.077)	(0.167)	(0.129)	(0.216)
4. Spouse has the same primary ancestry	-0.181	-0.091	-0.212	-0.144	-0.509 **
	(0.129)	(0.072)	(0.156)	(0.139)	(0.197)
Panel C: Spouse's Age and Education					
1. Spouse age	-0.901 **	-0.913 **	-1.100 **	-0.978 *	-1.493 **
	(0.383)	(0.376)	(0.459)	(0.497)	(0.729)
2. Spouse years of schooling	2.378 ***	2.320 ***	2.471 ***	2.301 ***	2.794 ***
	(0.298)	(0.261)	(0.444)	(0.385)	(0.594)
Panel D: Spouse's Labor Market Outcomes					
1. Spouse log(wages last year)	0.302 ***	0.279 ***	0.349 ***	0.236 ***	0.395 **
	(0.086)	(0.066)	(0.123)	(0.089)	(0.175)
2. Spouse worked last year	0.079 ***	0.070 ***	0.063 **	0.063 **	0.091 **
	(0.023)	(0.020)	(0.030)	(0.028)	(0.046)
3. Both worked last year	0.122 ***	0.116 ***	0.105 ***	0.102 ***	0.115 *
	(0.029)	(0.025)	(0.038)	(0.034)	(0.060)
Panel E: Fertility					
1. Number of children living in same household	-0.441 ***	-0.399 ***	-0.349 ***	-0.414 ***	-0.476 ***
	(0.075)	(0.065)	(0.089)	(0.087)	(0.160)
2. Has a child living in same household	-0.073 *	-0.041 *	-0.058	-0.058	-0.113 *
	(0.038)	(0.023)	(0.046)	(0.043)	(0.066)
3. Number of children living in same household, only individuals married with spouse present	-0.410 ***	-0.397 ***	-0.326 ***	-0.383 ***	-0.433 ***
	(0.058)	(0.053)	(0.087)	(0.067)	(0.147)
4. Has a child living in same household, only individuals married with spouse present	-0.007	0.004	-0.003	0.011	-0.008
	(0.020)	(0.018)	(0.033)	(0.018)	(0.046)
5. Is a single parent	0.021	0.034 *	0.025	0.033	0.010
	(0.025)	(0.020)	(0.031)	(0.026)	(0.045)
6. Is a never-married single parent	-0.002	0.008	0.001	0.005	-0.009
	(0.018)	(0.015)	(0.021)	(0.020)	(0.031)
Panel F: Residential Location					
1. Fraction of PUMA population from same country of birth	-0.007	0.0003	-0.014	-0.005	-0.027 **
	(0.008)	(0.003)	(0.012)	(0.008)	(0.012)
2. Fraction from same country of birth is above national mean for the country of birth	-0.015	0.022	-0.035	0.018	-0.189 ***
	(0.068)	(0.047)	(0.068)	(0.055)	(0.073)
3. Fraction of PUMA population with same primary ancestry	-0.018	-0.008	-0.032	-0.016	-0.064 ***
	(0.017)	(0.012)	(0.020)	(0.017)	(0.015)
4. Fraction with same ancestry primary is above national mean for the primary ancestry	-0.022	0.030	-0.032	-0.005	-0.188 ***
	(0.062)	(0.036)	(0.062)	(0.056)	(0.057)

Notes: The data and specification are the same as in Table 3, Column 2 for Panels A-D, Table 4, Column 2 for Panel E, and Table 5, Column 2 for Panel F with the following modifications: (1) Column 2 contains an additional control variable, max(0, age at arrival - 9)×log 1980 GDP of country of birth; (2) Column 3 contains an additional control variable, max(0, age at arrival - 9)×total fertility rate of country of birth in 1982; (3) Column 4 excludes people born in Canada; and (4) Column 5 excludes people born in Mexico. Each lettered row of each column is from a separate 2SLS regression that uses max(0, age at arrival - 9)×non-English-speaking country as the identifying instrument, is weighted by IPUMS weights and contains dummies for country of birth, age at arrival, age, race, Hispanic origin and sex. Robust standard errors adjusted for country of birth clusters are shown in parentheses. Asterisks denote significance levels (*=.10, **=.05, ***=.01).

Appendix Table 1. Individuals by Country of Birth

Panel A. English-speaking countries (=Control Group)

Rank by N	country	N	% of group
1	Canada	8,962	34.6%
2	England	6,121	23.6%
3	Jamaica	3,180	12.3%
4	United Kingdom, ns	1,242	4.8%
5	Trinidad & Tobago	1,014	3.9%
6	Guyana/British Guiana	991	3.8%
7	Scotland	803	3.1%
8	Ireland	565	2.2%
9	Australia	543	2.1%
10	South Africa (Union of)	308	1.2%
11	Barbados	297	1.1%
12	Bermuda	283	1.1%
13	Bahamas	258	1.0%
14	U.S. Virgin Islands	256	1.0%
15	Belize/British Honduras	251	1.0%
16	New Zealand	131	0.5%
17	Antigua-Barbuda	112	0.4%
18	St. Vincent	90	0.3%
19	Liberia	84	0.3%
20	Grenada	82	0.3%
21	St. Kitts-Nevis	73	0.3%
22	Wales	71	0.3%
23	Northern Ireland	69	0.3%
24	Zimbabwe	63	0.2%
25	St. Lucia	59	0.2%
26	British Virgin Islands	1	0.0%
27	Anguilla	1	0.0%
	Total English-spking obs	*25,910*	100.0%

Panel B. Non-English-speaking countries (=Treatment Group)

Rank by N	country	N	% of group
1	Mexico	47611	28.7%
2	Germany	19445	11.7%
3	Puerto Rico	13203	8.0%
4	Cuba	9389	5.7%
5	Vietnam	6334	3.8%
6	Italy	5642	3.4%
7	Japan	5475	3.3%
8	Korea	3926	2.4%
9	El Salvador	3233	2.0%
10	Dominican Republic	3103	1.9%
11	France	2466	1.5%
12	Portugal	2390	1.4%
13	Colombia	2266	1.4%
14	Taiwan	1987	1.2%
15	China	1854	1.1%
16	Laos	1668	1.0%
17	Poland	1499	0.9%
18	Haiti	1468	0.9%
19	Guatemala	1452	0.9%
20	Greece	1427	0.9%
21	Panama	1415	0.9%
22	South Korea	1344	0.8%
23	Ecuador	1316	0.8%
24	Iran	1314	0.8%
25	Spain	1207	0.7%
26	Netherlands	1188	0.7%
27	Nicaragua	1186	0.7%
28	Cambodia (Kampuchea)	1081	0.7%
29	Israel/Palestine	983	0.6%
30	Peru	933	0.6%
31	Argentina	926	0.6%
32	Thailand	818	0.5%
33	Honduras	801	0.5%
34	Austria	781	0.5%
35	Brazil	751	0.5%
36	Africa, ns/nec	680	0.4%
37	Venezuela	644	0.4%
38	Lebanon	637	0.4%
39	Hungary	550	0.3%
40	Turkey	544	0.3%
41	Azores	492	0.3%
42	Yugoslavia	487	0.3%
43	Costa Rica	466	0.3%
44	Chile	465	0.3%
45	Egypt/United Arab Rep.	454	0.3%
46	Iraq	443	0.3%
47	Other USSR/Russia	416	0.3%
48	Belgium	392	0.2%
49	Romania	358	0.2%
50	Indonesia	358	0.2%
	subtotal, top 50 countries	159,268	96.1%
	subtotal, other (91) countries	6,397	3.9%
	Total non-Eng-spking obs	*165,665*	100.0%

Notes: Information on each country's official languages is from the World Almanac. Recent adult immigrants from the 1980 IPUMS were used to divide English-official countries into English-speaking (at least 50% of recent adult immigrants did not speak a language other than English at home) or Other. The countries in the "Other" category are the Philippines, India, Hong Kong, Guam, Pakistan, Nigeria, American Samoa, Fiji, Tonga, Ghana, Kenya, Singapore, Dominica, Tanzania, Uganda, Sierra Leone, Senegal, Malta, Zambia, Micronesia, Marshall Islands, Papua New Guinea, Kiribati, Palau, Gambia, Malawi, Mauritius and Swaziland; people from these countries have been dropped from the empirical analysis.
Above tabulations by country of birth use following sample: individuals from the 2000 1% and 5% PUMS files who are currently aged 25-55, immigrated to the U.S. before age 15 and has nonmissing age, year of immigration, country of birth and English variables. Country refers to IPUMS detailed birthplace code.

Appendix Table 2. 2SLS Effect of English Using Alternative Instruments

	Base results (1)	Instrument is non-English × max(0, age at arrival - 7) (2)	Instrument is non-English × max(0, age at arrival -11) (3)	Instrument is non-English × age at arrival squared (4)	Instruments are non-English × age at arrival dummies (5)
Panel A: Marital Status					
1. Is currently married with spouse present	-0.108 ***	-0.126 ***	-0.089 **	-0.136 ***	-0.124 ***
	(0.040)	(0.043)	(0.043)	(0.044)	(0.043)
2. Is currently divorced	0.052 ***	0.055 ***	0.049 ***	0.061 ***	0.056 ***
	(0.016)	(0.018)	(0.016)	(0.021)	(0.019)
3. Has ever married	-0.072 **	-0.086 ***	-0.057 *	-0.088 ***	-0.082 ***
	(0.028)	(0.031)	(0.030)	(0.029)	(0.029)
Panel B: Spouse's Nativity and Ethnicity					
1. Spouse English-speaking ability ordinal measure	0.814 ***	0.849 ***	0.777 ***	0.881 ***	0.848 ***
	(0.072)	(0.082)	(0.058)	(0.085)	(0.078)
2. Spouse is US-born	0.322 **	0.381 ***	0.255 **	0.434 ***	0.377 ***
	(0.129)	(0.144)	(0.118)	(0.159)	(0.144)
3. Spouse has the same country of birth	-0.351 ***	-0.407 ***	-0.292 **	-0.457 ***	-0.403 ***
	(0.126)	(0.136)	(0.118)	(0.146)	(0.136)
4. Spouse has the same primary ancestry	-0.181	-0.232	-0.130	-0.274 *	-0.228
	(0.129)	(0.142)	(0.118)	(0.153)	(0.141)
Panel C: Spouse's Age and Education					
1. Spouse age	-0.901 **	-1.022 ***	-0.681	-0.956 ***	-0.924 ***
	(0.383)	(0.338)	(0.536)	(0.313)	(0.335)
2. Spouse years of schooling	2.378 ***	2.533 ***	2.251 ***	2.696 ***	2.535 ***
	(0.298)	(0.360)	(0.241)	(0.400)	(0.350)
Panel D: Spouse's Labor Market Outcomes					
1. Spouse log(wages last year)	0.302 ***	0.357 ***	0.280 ***	0.377 ***	0.337 ***
	(0.086)	(0.094)	(0.095)	(0.096)	(0.093)
2. Spouse worked last year	0.079 ***	0.077 ***	0.079 ***	0.080 ***	0.079 ***
	(0.023)	(0.024)	(0.029)	(0.022)	(0.023)
3. Both worked last year	0.122 ***	0.126 ***	0.112 ***	0.142 ***	0.129 ***
	(0.029)	(0.029)	(0.034)	(0.029)	(0.029)
Panel E: Fertility					
1. Number of children living in same household	-0.441 ***	-0.465 ***	-0.408 ***	-0.485 ***	-0.465 ***
	(0.075)	(0.080)	(0.082)	(0.087)	(0.082)
2. Has a child living in same household	-0.073 *	-0.086 **	-0.060	-0.094 **	-0.085 **
	(0.038)	(0.041)	(0.037)	(0.044)	(0.041)
3. Number of children living in same household, only individuals married with spouse present	-0.410 ***	-0.417 ***	-0.388 ***	-0.440 ***	-0.423 ***
	(0.058)	(0.058)	(0.063)	(0.071)	(0.062)
4. Has a child living in same household, only individuals married with spouse present	-0.007	-0.013	0.002	-0.020	-0.013
	(0.020)	(0.022)	(0.019)	(0.028)	(0.023)
5. Is a single parent	0.021	0.023	0.015	0.024	0.022
	(0.025)	(0.026)	(0.024)	(0.025)	(0.025)
6. Is a never-married single parent	-0.002	-0.0001	-0.007	-0.001	-0.001
	(0.018)	(0.018)	(0.019)	(0.018)	(0.018)
Panel F: Residential Location					
1. Fraction of PUMA population from same country of birth	-0.007	-0.010	-0.005	-0.0119	-0.010
	(0.008)	(0.009)	(0.008)	(0.010)	(0.009)
2. Fraction from same country of birth is above national mean for the country of birth	-0.015	-0.028	7.95E-05	-0.040	-0.028
	(0.068)	(0.078)	(0.063)	(0.088)	(0.079)
3. Fraction of PUMA population with same primary ancestry	-0.018	-0.025	-0.012	-0.032	-0.025
	(0.017)	(0.019)	(0.016)	(0.021)	(0.019)
4. Fraction with same ancestry primary is above national mean for the primary ancestry	-0.022	-0.040	-0.007	-0.046	-0.035
	(0.062)	(0.071)	(0.059)	(0.077)	(0.071)

Notes: Column 1 is identical to Table 6, Column 1, and shows the original 2SLS estimates of the effect of English in which the identifying instrument is max(0, age at arrival - 9) × non-English-speaking country. The identifying instrument differs in Columns 2-5 as follows: it is max(0, age at arrival - 7)× non-English-speaking countr in Column 2; max(0, age at arrival - 11)× non-English-speaking country in Column 3; age at arrival squared× non-English-speaking country in Column 4; and (4) a full set of age at arrival dummies interacted with non-English-speaking country in Column 5. Each lettered row of each column is from a separate 2SLS regression that uses the aforementioned identifying instrument, is weighted by IPUMS weights and contains dummies for country of birth, age at arrival, age, race, Hispanic origin and sex. In addition, in Column 4, we also control for age at arrival squared. Robust standard errors adjusted for country of birth clusters are shown in parentheses. Asterisks denote significance levels (*=.10, **=.05, ***=.01).