DOCUMENT RESUME

ED 473 464 UD 035 492

AUTHOR	Bleakley, Hoyt; Chin, Aimee
TITLE	Language Skills and Earnings: Evidence from Childhood Immigrants.
PUB DATE	2002-10-01
NOTE	54p.
PUB TYPE	Reports - Research (143)
EDRS PRICE	EDRS Price MF01/PC03 Plus Postage.
DESCRIPTORS	Age Differences; *Educational Attainment; *English (Second Language); *Immigrants; *Language Skills; *Wages; Young Children

ABSTRACT

Because research on the effect of language skills on earnings is complicated by the endogeneity of language skills, this study exploited the phenomenon that younger children learn languages more easily than older children to construct an instrumental variable for language proficiency. Data came from the 1990 U.S. Census, specifically the Integrated Public Use Microsample Series files. The sample included people arriving between 1960-74 who had been in the United States for 16-20 years and were age 25-38 in 1990. The sample was divided into three mutually exclusive language categories: non-English-speaking countries of birth, countries of birth with English as an official language that had English as the predominant language, and other countries of birth with English as an official language. Results found a significant positive effect of English proficiency on wages among adults who immigrated to the United States as children. Much of this impact appeared to be mediated through education. Differences between non-English-speaking origin countries and English-speaking ones that might make immigrants from the latter a poor control group for non-language age-at-arrival effects did not drive these findings. (Contains 33 references.) (SM)

LANGUAGE SKILLS AND EARNINGS: EVIDENCE FROM CHILDHOOD IMMIGRANTS*

Hoyt Bleakley
University of Chicago (NORC)

Aimee Chin
University of Houston

This version: October 1, 2002

* Bleakley: Postdoctoral Fellow, University of Chicago (NORC), 1155 East 60th Street, Chicago, IL 60637 (email: bleakley@uchicago.edu); Chin: Assistant Professor, Department of Economics, University of Houston, 204 McElhinney Hall, Houston, TX 77204-5019 (email: achin@uh.edu). We thank Daron Acemoglu, Josh Angrist, David Autor, Esther Duflo, Amy Finkelstein, Mark Lewis, Robin McKnight, Bob Triest, and participants in the MIT labor lunch, MIT labor/public economics seminar, Rice/UH empirical economics seminar and NBER Summer Institute for helpful comments and discussion. The authors bear sole responsibility for the content of this paper.

LANGUAGE SKILLS AND EARNINGS: EVIDENCE FROM CHILDHOOD IMMIGRANTS

ABSTRACT

Research on the effect of language skills on earnings is complicated by the endogeneity of language skills. This study exploits the phenomenon that younger children learn languages more easily than older children to construct an instrumental variable for language proficiency. We find a significant positive effect of English proficiency on wages among adults who immigrated to the U.S. as children. Much of this impact appears to be mediated through education. Differences between non-English-speaking origin countries and English-speaking ones that might make immigrants from the latter a poor control group for non-language age-at-arrival effects do not drive these findings. (JEL J61, J24, J31)

I. Introduction

For both social and economic reasons, language is a barrier that separates many immigrants from natives. On the social side, immigrants who speak English poorly are more visibly foreign than others. This may facilitate discrimination on the part of natives, and contribute to social isolation and ghettoization that makes immigrants feel less American. On the economic side, weak language skills probably reduce productivity and therefore increase the immigrant-native earnings gap. Moreover, strong language skills almost certainly increase the range and quality of jobs that immigrants are likely to get. This view is supported by numerous empirical studies which suggest a positive association between English-language ability and earnings.[1]

Interest in the language skills of immigrants has been fostered in part by the recent upsurge in immigration to the United States. The 2000 Census showed that 11 percent of the U.S. population is foreign born, up from 8 percent in 1990. Most of these recent immigrants are from non-English-speaking countries. In fact, the 2000 Census also showed that 45 million U.S. residents age 5 and over spoke a language other than English at home and 20 million spoke English less than "very well".

Although language is central to the process of immigrant assimilation, and the relationship between language and earnings has been the subject of considerable research, the problem of measuring the causal effect of language skills on earnings is complicated by the fact that immigrants with stronger language skills may earn more for reasons other than these skills. Studies to date have relied primarily on simple regression strategies to control for confounding factors.

[1] See Section II for an overview of these studies.

The contribution of this paper is the implementation of an identification strategy for the causal effect of language skills that is motivated by research on language acquisition. Younger children tend to learn languages easily while adolescents and adults do not. This psychobiological phenomenon leads us to use an instrumental variable derived from immigrants' age at arrival. As we show below, there is a powerful association between immigrants' age at arrival and language skills in the 1990 Census. On the other hand, age at arrival probably affects immigrant earnings through channels other than language. For example, immigrants who arrive earlier may adapt more quickly to American institutions. We therefore use immigrants from English-speaking countries to control for secular (*i.e.*, non-language-related) age-at-arrival effects. The result is an instrumental variable (IV) strategy using age at arrival interacted with a dummy for non-English-speaking country as the identifying instrument.

To make this idea concrete, consider four immigrants, each brought to the U.S. as a child. Two are from Jamaica (an English-speaking country), one aged five at arrival and the other aged fifteen. The other two are from Mexico (a non-English-speaking country), with parallel ages of arrival. If we observe a difference between the wages of the two Jamaicans, we could attribute it to secular age-at-arrival effects. But all of these effects are also present in the case of the two Mexicans, *in addition* to the fact that the Mexicans had substantially less exposure to the English language before immigrating. As such, the Jamaicans can be used to control for the secular age-at-arrival effects. Any differences between the Mexicans in excess of the differences between the Jamaicans can be attributed to *language* effects, *i.e.*, that the child who immigrated to the U.S. at an older age had a higher cost of acquiring a second language, and thus attained a lower level of proficiency in English.

Using individual-level data from the 1990 U.S. Census, we find that English-language skills have substantial, positive effects on wages and educational attainment. The IV estimates are higher than the ordinary least squares (OLS) estimates; the latter are subject to upward bias resulting from ability bias that is obscured by severe downward bias resulting from measurement error in the language skills variable. Most of the effect of language skills on wages appears to be mediated by the effect on years of schooling. This suggests that the role of language proficiency as an input to the production of human capital is far more important than the direct effect of language on the marginal product of labor.

One important concern regarding the interpretation of our results is whether immigrants from English-speaking countries provide a good control for secular age-at-arrival effects. Considering that English-speaking countries tend to be richer than non-English-speaking countries, there might be concomitant differences that affect an immigrant's progress in the U.S. To enhance comparability between the treatment and control countries, we incorporate country-of-birth school-quality variables into the regressions. In particular, we allow these variables to independently shift the age-at-arrival effects on language, wages, and education. Doing so does not affect our results.

A second, closely related, concern is that our sample is dominated by Mexicans and Canadians. While it might be reasonable to argue that immigrants from English- and non-English-speaking countries experience the same non-language age-at-arrival effects where Mexicans and Jamaicans are concerned, this argument appears tenuous for Mexicans and Canadians. Since Canadians likely have a shorter "cultural distance" from Americans, they should have lower age-at arrival effects than Mexicans, such that the causal effects of language skills that we estimate would be upward biased. In view of this concern, we perform robustness checks in which we drop individual countries or groups of countries. All our results remain, albeit with higher standard errors, including when the analysis excludes both Mexicans and Canadians, as well as when it is includes only Caribbean nations.

The rest of the paper is organized as follows. Section II discusses the literature on the returns to language skills on the one hand and language acquisition on the other, and describes the data used in our empirical analysis. Section III presents the base results. Section IV performs some robustness checks and discusses some implications of the findings. Section V concludes.

II. Background and Data

A. Previous Research on Language Skills and Earnings

This study has several antecedents in the literature. One set of studies focuses on how long it takes for immigrant workers to achieve earnings parity with native-born workers (see Schultz (1998) and Borjas (1999) for reviews; also Friedberg (1993, 2000)). Their finding of an initial earnings disadvantage for immigrants that decreases with years in the host country is certainly consistent with the language skills hypothesis; however it is also consistent with numerous other explanations.

A second, related set of studies seeks to explicitly test the language skills hypothesis. Earlier studies tend to regress log earnings on some measure of language skills and interpret the OLS coefficient for the language variable as the labor market return to language skills (*e.g.*, McManus, Gould and Welch (1983), Kassoudji (1988), Tanier (1988) and Chiswick (1991)). More recent studies have attempted to address the problem of endogeneity in the relationship between language and earnings (*e.g.*, Chiswick and Miller (1992, 1995, 1999), Angrist and Lavy (1997), and Dustmann and van Soest (2002)).

Angrist and Lavy use an IV strategy based on a policy change in the schooling system of Morocco. However, the context of their "natural experiment" is quite different from ours: they estimate the return to speaking French in Morocco, an Arabic-speaking country, among native Moroccans. It is unclear that the lessons learned in their study can be readily extrapolated to the situation of immigrants in the U.S. labor market.[2]

Dustmann and van Soest (2002) and the Chiswick and Miller studies analyze the returns to proficiency in the dominant language. Chiswick and Miller's identifying instruments include minority-language concentration of the place of residence, veteran status, whether married overseas and number of children. However, the excludability of their instruments from the wage equation has been called into question (Borjas (1994)).[3] Dustmann and van Soest address the problem of time-invariant unobserved individual heterogeneity by using fixed effects estimation (they use panel data for West German immigrants extracted from the Germany Socioeconomic Panel). In addition, to approach the potential problems of time-varying unobserved individual heterogeneity and measurement error in the language proficiency measure, they use instrumental variables. Some of their identifying instruments, such as parents' education, are subject to the caveats mentioned for the Chiswick and Miller studies.

A third set of studies has documented the low educational attainment among childhood immigrants. Individuals who immigrated from Mexico and Central America as children are much less likely than natives to complete high school and indeed even junior high school

[2] French is not the predominant language of Morocco, although as a vestige of the country's colonial history it continues to be used in the civil service and trade-oriented sectors. On the other hand, English is the dominant language of the U.S., and the lack of English-language skills impedes participation in a much broader range of jobs and sectors.

[3] For example, the concentration ratio is a region-of-residence variable, but region of residence is a choice variable, and regions with higher concentrations differ from regions with lower concentrations in a variety of ways, one of which is language. Moreover, regional characteristics correlated with the concentration ratio (*e.g.*, industrial composition, extent of ethnic businesses, extent of poverty) have direct effects on earnings. In general, one's region of residence, household composition, human capital investment and labor market decisions are jointly determined, *i.e.*, all outcomes of the same household utility maximization problem.

(Hispanic Dropout Project (1998) and Urban Institute (2000)). We are unaware of studies that rigorously identify the determinants of the immigrant-native gap in educational attainment. Furthermore, we believe that the present study is the first to identify the contribution of language proficiency to earnings through pre-market factors such as education.

B. Language Acquisition Theory and Empirical Research

Our choice of instrument is motivated by the well-documented relationship between language acquisition and age in the psychobiological literature. Younger children learn languages more easily than adolescents and adults. Cognitive scientists refer to this as the "Critical Period Hypothesis". There is believed to be a critical age range in which individuals learn languages more easily and after which language acquisition is more difficult. If exposure to the language begins during the critical period, acquisition of the language up to "native" ability is almost automatic. If exposed afterwards, the individual's performance is less certain.

Behavioral evidence has been supportive of this hypothesis: late learners tend to attain a lower level of language proficiency (see Newport (1990) for a review). This appears to be linked to physiological changes in the brain (Lenneberg (1967)). Maturational changes starting just before puberty precipitate a sharp reduction in a child's ability to acquire second languages, especially with respect to sound production and grammatical structure, and to lesser extent vocabulary.

Applied to immigrants to the U.S., the Critical Period Hypothesis predicts that those who arrive at an earlier age will develop better English-language skills than those who arrive at a later age. We test this prediction after describing our data.

C. Data and Descriptive Statistics

We implement our empirical strategy using microdata from the 1990 U.S. Census, specifically the Integrated Public Use Microsample Series (IPUMS) files (Ruggles, *et al.*

(1997)). We combine the 5 percent State sample with the 1 percent Metro sample.[4] We restrict our attention to childhood immigrants, which we define as those immigrants who were under age 18 upon arrival to the U.S. For these individuals, age at arrival is not a choice variable since they did not time their own immigration but merely followed their parents to the U.S.[5] Year of arrival to the U.S. is reported in multi-year intervals, with more detailed intervals for the recent past.[6] Our definition of age at arrival is [current age – (1990 – maximum year of arrival)], so we are using the maximum possible age at arrival. We choose this conservative definition of age at arrival so as not to mistakenly include adult migrants in our sample.

We also impose the following restrictions. First, they arrived between 1960 and 1974, or equivalently, they have been in the U.S. for 16 to 30 years. Second, they are between age 25 and 38 in 1990. The first cutoff selects individuals who would have likely completed schooling. The second cutoff is a result of our age at arrival and year of arrival restrictions. Our results do not change qualitatively when any of these cutoffs are changed.

We divide our sample into three mutually exclusive language categories: 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.[7] The first category is our "treatment" group and the second is our "control" group.

[4] The weights are adjusted to reflect the fact that the Metro sample is one-fifth of the State sample.

[5] According to the U.S. Immigration and Naturalization Service, immigrating parents may bring any unmarried children under age 21. This paper uses a more restricted set of childhood immigrants: immigrants who were under 18 upon arrival (i.e., maximum age at arrival is 17).

[6] Year of arrival to the U.S. data is reported in intervals, i.e., before 1950, 1950-1959, 1960-1964, 1965-1969, 1970-1974, 1975-1979, 1980-1981, 1982-1984, 1985-1986 and 1987-1990.

[7] We used *The World Almanac and Book of Facts, 1999*, to determine whether English was an official language of

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.[8] Appendix Table 1 displays the categorization of countries, as well as the composition of our sample by national origin.

Table 1 provides the descriptive statistics for the treatment and control groups. They are separately reported for those immigrants arriving at a younger age (0 to 11) and older age (12 to 17). English-speaking ability[9] is higher for younger arrivers from non-English-speaking

---

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.

[8] Our results do not change when we include these omitted English-official countries. Because this group has had some intermediate level of exposure to English prior to arrival, when we estimate the regressions in Section III using it as the control and using the non-English-speaking countries as the treatment, the first stage and reduced-form coefficients are lower in magnitude, but the 2SLS coefficients are about the same.

[9] 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." Other studies have used this question to study English proficiency, and have likewise coded immigrants who speak only English as speaking English very well (*e.g.*, Chiswick and Miller (1992, 1995)). The English-speaking ability measure is coded as 0 for not speaking English at all, 1 for speaking English not well, 2 for speaking English well and 3 for speaking English very well.

countries, but not different for young arrivers from English-speaking countries. The ordinal measure of English-speaking ability is higher for younger arrivers from non-English-speaking countries but similar across age-at-arrival categories for immigrants from English-speaking countries. Wages[10] are not different for younger arrivers from non-English-speaking countries, but lower for younger arrivers from English-speaking countries. This latter observation reflects the upward sloping relationship between age and wage (young arrivers are on average four years younger than older arrivers); interestingly, this relationship is not borne out among immigrants from non-English-speaking countries. Years of schooling are higher for immigrants from English-speaking countries, and for younger arrivers. Immigrants from non-English-speaking countries are more likely to be Hispanic whereas those from English-speaking countries are more likely to be white or black.

III. Estimation Results

A. Reduced-form Estimation

Simple statistical techniques can be used to illustrate how the IV strategy based on age at arrival identifies the effect of English-language skills on wages. Consider the regression model,

$$(1) \qquad y_{ija} = \alpha + \beta x_{ija} + \delta A_a + \gamma N_j + \varepsilon_{ija}$$

for individual i born in country j arriving to the U.S. at age a. y_{ija} is log wages, x_{ija} is a measure of English-language skills (the endogenous regressor), A_a is a dummy for arrived young (age at arrival ≤ 11) and N_j is a dummy for born in a non-English-speaking country. Let z_{ija} denote the binary instrument, the interaction between arrived young and born in a non-English-speaking country, *i.e.*, $z_{ija} = A_a{*}N_j$. The IV estimate of β in this equation is

[10] We only use individual's income from wage and salary because we are interested in estimating the labor market return to English-language skills.

$$(2) \qquad \beta_{IV} = \frac{(\bar{y}_{1,1} - \bar{y}_{0,0}) - (\bar{y}_{1,0} - \bar{y}_{0,1})}{(\bar{x}_{1,1} - \bar{x}_{0,0}) - (\bar{x}_{1,0} - \bar{x}_{0,1})}$$

where $\bar{y}_{1,0}$ is the mean of y_{ija} for those observations with $A_a = 1$ and $N_j = 0$; other terms are similarly defined. The numerator is the reduced-form relationship between y_{ija} and z_{ija}: the difference-in-difference of mean log earnings. The denominator is the reduced-form relationship between x_{ija} and z_{ija}: the difference-in-difference of mean English ability. The β_{IV} obtained from estimating Equation 1 using two-stage least squares (2SLS) is identical to the indirect least squares estimate obtained from taking the ratio of the reduced-form coefficients since Equation 1 is just-identified.

We emphasize that the identifying instrument is not age at arrival itself. The latter exclusion restriction seems difficult to justify *a priori*, since younger arrivers likely differ from older arrivers along non-language dimensions that also affect earnings. For example, in addition to having earlier exposure to English, younger arrivers are matriculated into the U.S. educational system at an earlier age. To the extent that human capital acquired in U.S. schools is better suited to the U.S. labor market, the younger arrivers would have an advantage that has nothing to do with language human capital (Friedberg (2000)). Also, younger children may face lower costs of assimilation along cultural dimensions that also have nothing to do with language *per se*. Furthermore, families that migrate with younger children may differ along some important margin from those that migrate with older children.

Instead, the identifying instrument is an interaction of age at arrival with country of birth. 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 to 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 wages between young and old arrivers in non-English-speaking countries that is over and above the difference in English-speaking countries can plausibly be attributed to language.

The relationship between age at arrival and English-language skills is shown graphically in Figure 1. The diamond-marker line in Panel A displays the mean English-speaking ability for immigrants from non-English-speaking countries. Consistent with the research on language acquisition, children who received their first exposure to English at an earlier age attain a higher level of English-language proficiency than those who received it later. In fact, immigrants from non-English-speaking countries who arrive quite young (up until age 8 or 9) attain English-language skills comparable to those of immigrants from English-speaking countries. After that age, however, their English-language skills decline markedly, with older arrivers attaining progressively lower proficiency.

The square-marker line in Panel A displays the mean English-speaking ability of the immigrants from English-speaking countries. It is flat: nearly every immigrant from English-speaking countries speaks English very well.[11] This is not surprising: their first exposure to English does not depend on when they migrated to the U.S. This supports our assertion that the pattern for immigrants from non-English-speaking countries is related to second language acquisition, and not to some spurious relationship in our sample between age at arrival and English-speaking ability.

Figure 1, Panel B displays the difference in mean English-speaking ability between immigrants from English- and non-English-speaking countries. Older arrivers have statistically significantly lower English-speaking ability. This same result is summarized in Table 2. Early arrivers are 1.42 percent more likely to speak at least some English (Column 2), 7.94 percent

[11] This line is not mechanically pinned at three because some of these countries have large non-English-speaking communities, e.g., the Quebecois in Canada.

more likely to speak English well or very well (Column 4), and 21.88 percent more likely to speak English very well (Column 6). These increases at each point in the cumulative distribution function (CDF) of English-speaking ability translate into increases in the mean of the ordinal measure of English-speaking ability: the ordinal measure is 0.3124 units higher for early arrivers (Column 8).

Figure 2 shows the relationship between age at arrival and wages. The similarity to Figure 1 is striking. Panel A shows the mean log annual wages as a function of age at arrival for immigrants from non-English-speaking countries and for those from English-speaking countries. As in Figure 1, Panel A, the lines corresponding to the means of the two groups are similar at earlier ages at arrival and diverge for later ages. Among the younger arrivers, whether they come from non-English-speaking countries makes no significant difference in their wages. Among the adolescent arrivers, however, wages tend to be lower for the immigrants from non-English-speaking countries. The line for immigrants from English-speaking countries is nearly flat, suggesting that the non-language effects of age at arrival are small.[12] Panel B shows the difference in mean between the two groups. The differential drop in wages for older arrivers closely parallels the differential drop in English-speaking ability for older arrivers shown in Figure 1, Panel B.

The information contained in Figures 1 and 2 can be used to construct the indirect least squares estimate given in Equation 2. The numerator would be derived from Figure 2, Panel B: calculate the mean difference in means for each the younger arrivers (0 to 11) and the older

[12] Alternatively, this might suggest that immigrants from English-speaking countries are a poor control group, since they do not capture all the non-language age-at-arrival effects that immigrants from non-English-speaking countries experience. In Section IV, we will attempt to enhance comparability between English- and non-English-speaking countries in a variety of ways.

arrivers (12 to 17), and then take the difference. The denominator would be similarly derived from Figure 1, Panel B. This exercise is equivalent to taking the ratio of the reduced-form coefficient from a regression with y_{ija} as the left-hand-side variable and the reduced-form coefficient with x_{ija} as the left-hand-side variable. These reduced-form coefficients are shown in Table 2. We obtain an indirect least squares estimate of returns to each unit of English-speaking ability of 39 log points.[13] That is, one additional unit of English-speaking ability raises wages by about 39 percent. This compares with an OLS estimate of 22 percent (from Table 3 to be discussed below). Thus, the IV estimate suggests that the OLS estimate is downward biased.

The "arrived young" main effect is consistently positive in Table 2 (even columns). This suggests that simple-difference estimates with just immigrants from non-English-speaking countries (instead of difference-in-differences estimates with immigrants from English-speaking countries also) would have overstated the effect of English-language skills by neglecting secular age-at-arrival effects. However, this effect is substantially smaller than the estimated effect of age at arrival for immigrants who originated from non-English-speaking countries. Additionally, the "non-English speaking country of birth" main effect is consistently negative, which is as expected: childhood immigrants originating from English-speaking-countries on average attain a higher level of English-language proficiency as adults.

Investment in education may be an important intervening factor in the effect of language skills on earnings, as suggested by Figure 3. The pattern of years of schooling completed by age at arrival bears remarkable resemblance to the pattern of earnings by age at arrival. In examining the economic returns to language skills, therefore, it is essential to recognize that language can affect earnings through direct as well as indirect channels.

[13] Numerator is from Column 10: 0.1221. Denominator is from Column 8: 0.3124.

B. Two-Stage Least Squares Estimation

The previous subsection made simplifications to illustrate the IV strategy. In this subsection, we drop the assumption that age at arrival is binary, and proceed to use age at arrival in a way that better captures the pattern of second-language acquisition in children. We use a parameterization that admits a degradation in language-learning ability that starts at age twelve and grows linearly: max(0, $a_i - 11$), in which a_i continues to be individual i's age at arrival. Of course, the key prediction is that the immigrants from English- and non-English-speaking countries have increasingly divergent language and wage outcomes starting at age-at-arrival twelve, so the instrument excluded from the second stage is $k_{ija} = \max(0, a_i - 11)*N_j$.[14] This piecewise-linear variable allows the difference between the control (English-speaking country of birth) and treatment (non-English-speaking country of birth) groups to grow starting just before the onset of puberty.

The above procedure is summarized by the following two-equation system. The second-stage equation relates the outcome of interest, wages, to the endogenous regressor, English-language skills. This is just Equation 1, which is modified here by the inclusion of a vector of exogenous covariates $\mathbf{w}_{ija}$:

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

The first-stage equation relates the endogenous regressor to the instrument k_{ija}:

(4) $x_{ija} = \alpha_1 + \beta_1 k_{ija} + \delta_{1a} + \gamma_{1j} + \rho_1'\mathbf{w}_{ija} + \varepsilon_{1ija}$.

This system is just-identified. δ_a is a full set of age-at-arrival fixed effects; this controls for non-language age-at-arrival effects in a finer way than just having a dummy for arriving young. γ_j is

[14] Results are not dependent on our particular parameterization of age at arrival. Appendix Table 2 presents results using alternative ways of defining the instrument.

a full set of country-of-birth fixed effects; this controls for differences in mean immigrant "quality" as reflected in wages from country to country more precisely than just having a dummy for originating from a non-English-speaking country.

1. Effect of language skills on earnings

The first-stage regression results (from estimating Equation 4) are displayed in Table 3, Columns 1 and 2. There is a strong, negative relationship between the instrument k_{ija} and English-speaking ability. Immigrants who arrived from non-English-speaking countries have progressively poorer English skills for each year of arrival past age 11. Even-numbered columns include controls for a full set of country of birth dummies while odd-numbered columns do not.

The results from estimating Equation 3 are displayed in the last four columns of Table 3; Columns 5 and 6 show the results using OLS and Columns 7 and 8 show the results using 2SLS. Column 8 suggests that on average, improving English-speaking ability by one unit increases wages by 33.35 percent. This 2SLS estimate of the return to one unit of English-speaking ability is higher than its OLS counterpart (22.19 percent in Column 6). The OLS estimate appears to be downward biased, although it should be noted that its 95 percent confidence interval overlaps with the 95 percent confidence interval of the 2SLS estimate. This is nevertheless somewhat surprising, since the ability bias story implies higher OLS estimates than IV estimates; this issue is discussed in Section IV.C.

At this point, it is worth pointing out that these results are robust to the exclusion of immigrants from Mexico or Canada, as shown in Table 4, left side. Excluding Mexicans results in the loss of over ten thousand observations, which is more than one-fourth of all immigrants from non-English-speaking countries in our sample. Excluding Canadians results in the loss of over three thousand observations, which is two-fifths of all immigrants from English-speaking countries. It is not surprising, therefore, that the standard errors are much larger. However, it may surprise some skeptics of our identification strategy that the magnitude of the 2SLS estimates is unchanged. If our base results were driven by a comparison between Mexicans and Canadians, then we should have obtained lower estimated returns to language when Mexico and Canada were dropped from the sample. This is because, as the story goes, Canadians are poor controls for the non-language age-at-arrival effects experienced by Mexicans; even if geographic distance is not different between the two, yet Canadians might be more culturally similar to Americans such that they may not be as sidetracked by a later age at arrival. This story does not appear to hold in our data, lending support to our difference-in-differences identification strategy and our interpretation of the 2SLS estimate as the return to language. We defer presenting additional robustness checks until Section IV.A.

2. Effect of language skills on educational attainment

Figure 3 had suggested that much of the effect of language skills on earnings could be channeled through investments in the education form of human capital. Since instruction in U.S. classrooms is almost exclusively conducted in English, English-language skills can be expected to affect not only the quality of learning at each stage of schooling and but also the probability of progression to the next stage of schooling. Individuals who have poorer English-language skills effectively face a higher cost of education – it may be impossible to master the materials, or at the very least it requires more effort to do so.

The OLS estimate of the effect of English-language skills on educational attainment might be biased for the same reasons that the OLS estimate of their effect on wages might be biased (*e.g.*, ability bias, measurement error, reverse causality). By using the exogenous variation provided by language-learning theory, we can obtain a consistent estimate of the effect of English-language skills on educational attainment. Table 4, right side displays the estimation results. We have estimated the models described by Equations 3 and 4 with years of schooling as the outcome of interest. The OLS estimate (Column 3) suggests that increasing English-speaking ability by one unit raises years of completed schooling by two years. The 2SLS estimate (Column 4) is twice the OLS estimate: on average, a one unit increase English-speaking ability raises educational attainment by four years.[15]

Besides affecting the mean years of schooling completed, language proficiency also appears to affect the distribution of educational attainment in the population. This is apparent in Figure 4, where we plot the probability distribution functions (PDFs) of educational attainment for two categories of age (young and old) and two categories of countries (non-English and English-speaking). The difference-in-differences in PDF is plotted in Panel D. Each point on the bold line comes from a separate regression with the probability of attaining a certain level of education as the left-hand-side variable and age, race/ethnicity and female dummies as additional controls. The graph shows a negative area for grades 5 to 11, indicating that poor English proficiency increased drop-out rates at these levels. The positive area from 12 to 15 suggests that better English-language skills increased the share of immigrants completing high school and attending some college. Better English-language skills do not appear to have changed the share of immigrants at the lowest and highest levels of education as much.

The results for education are quite striking: because they are assigned a higher cost of language acquisition, childhood immigrants who arrive to the U.S. at a later age are much less likely to either enter or graduate high school. This effect is so large that it may set off a few alarm bells. In particular, it could be indicative of dynamic differences between the treatment and control groups. For example, many low-educated young men migrate on their own to the U.S. from Mexico and Central America to look for work. These "loner" immigrants will almost

[15] This result is robust to excluding the oldest arrivals (ages 14-17) from the regression, as discussed below.

all enter the older children category (arrived > age 11), making older children systematically different from younger children. In particular, among the older children, there is a disproportionate number of low-educated immigrants who never intended to attend school in the U.S., and moreover who likely differ along other dimensions as well since they did chose to migrate on their own.[16] Our identification strategy is partly predicated on childhood immigrants being brought to the U.S. by a decision of their parents. Labeling these loner immigrants as children under our *gringo* definition of adulthood (i.e., eighteen and over) may be misleading.

To address the problem of the loner immigrants, we restrict our analysis to those who arrived to the U.S. at age fourteen or younger, *i.e.*, we drop the fifteen to seventeen-year-olds. Our results are qualitatively similar, although the point estimate is smaller (instead of a one-unit increase in English proficiency raising years of schooling by 4.2 years, it is now 3.3 years) and the standard errors are larger (since there are ten thousand fewer observations are lost). This suggests that what we observe is truly an effect of language and not due to the independent (and therefore possibly self-selected) migration of young adults.

IV. Some Specification Issues

In this section, we discuss the interpretation of our findings. Section A addresses the concern that the differential age-at-arrival effects for non-English-speaking countries may not be due to language, but some omitted factor that co-varies with age at arrival in the same way. Our findings survive a variety of robustness checks. We proceed in Section B to discuss the role of investments in education human capital in the effect of language proficiency on wages. Finally, Section C analyzes the role of measurement error in explaining the "puzzle" of why the IV estimates are higher than the OLS estimates of the return to language skills.

[16] Their uncertain immigration status and lack of access to capital markets may preclude enrollment anyway.

A. Additional Robustness Checks

We have been interpreting 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-language 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-language proficiency is invalid. In this subsection, we consider two hypotheses for differential age-at-arrival effects between the two groups of immigrants that have nothing to do with the causal effect of language skills. One alternative hypothesis is 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. Another alternative hypothesis is that parents from non-English-speaking countries may factor their children's ages into the migration decision in a way that is different from parents from English-speaking countries.

To preview the results, we find that even after allowing for differential age-at-arrival effects between poorer and richer countries, the estimates of the effect of each unit of English-speaking ability on wages remain around 30 percent. Additionally, there is no evidence that the age-at-arrival distribution is different between immigrants from English- and non-English-speaking countries, thus casting doubt on the second alternative hypothesis. These results therefore strengthen the case for interpreting the 2SLS estimate as the causal effect of English-language skills.

1. How comparable are treatment and control countries?

The first alternative hypothesis is 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, as seen in Appendix Table 1. 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 may reflect not only differential English-language skills but also differential returns to origin-country schooling.[17]

In Section III.B.1, we showed that our results were not sensitive to the exclusion of Mexicans and Canadians, which already provides some degree of assurance that our results are not driven by differential age-at-arrival effects between English- and non-English-speaking countries. To further assess this hypothesis, we adopt two tactics. First, we control explicitly for characteristics of the country of birth in the regression models. The country data that we employ are GDP, per pupil school expenditures and teacher-pupil ratio. We use the 1965 level of these characteristics, merged in from the Barro-Lee and Summers-Heston cross-country panel data sets. These variables should be correlated with the school quality prevailing in the country of birth. Including these characteristics as regressors would be useless: the country-of-birth fixed effects fully absorb them. Instead, we use the interactions between these characteristics and age at arrival. We do this because the value in the U.S. of schooling obtained in higher-school-quality countries may differ from schooling obtained in lower-school-quality countries. Since age at arrival affects the share of schooling obtained in the country of birth, the estimates of β above may reflect not only language effects, but also non-language effects (specifically,

[17] Immigrants who arrived at a younger age systematically receive a lower share of their schooling in their origin country. Friedberg (2000) finds that, among immigrants to Israel, there is a lower return to schooling obtained abroad than to schooling obtained in Israel. This, in and of itself, provides a strong additional justification for including a main effect of age at arrival. However, for this to impact our strategy, the effect has to vary between the control and treatment groups.

differential school quality). By controlling for the school quality interactions with age at arrival, we should purge the difference-in-differences of some of this non-language effect.

Table 5 shows the estimation results from adding these school-quality interactions one by one, and finally all three at once. The principal finding is that although the school quality interactions enter significantly in the first stage and reduced-form equations, the coefficient for k_{ija} remains negative and significant. The 2SLS estimates of the return to English-speaking ability remain around 30 percent. (We perform the same analysis with years of schooling instead of earnings as the outcome of interest, and the base result reported in Section III.B.2 that each unit improvement in English-speaking ability raises schooling by four years remains.)

The second tactic we take to assess the first alternative hypothesis is to match countries in the control group to countries in the treatment group to make them more comparable by such attributes as geography, history, and GDP. An advantage of this matching strategy is that it potentially controls for effects of country of birth characteristics that are nonlinear; the previous strategy of adding interactions between country of birth characteristics and age at arrival assumes that those characteristics have linear effects.[18] A limitation of this matching strategy is that degrees of freedom are drastically reduced.

Table 6 allows for different age-at-arrival effects between richer and poorer countries. Specifically, we allow the treatment effect and, in some specifications, the effect of the control variables to differ between immigrants from countries with below-median GDP and immigrants from countries with above-median GDP. The first stage results in Column 1 indicate that the instrument has a weaker effect on immigrants from richer countries. Additionally, the reduced-

[18] An example of a nonlinear effect might be that only if a country is beyond some threshold GDP does age at arrival cease to have an effect; it is not the case that for each additional dollar of GDP, age at arrival has a marginally smaller effect.

form effects on wages (presented in Column 2) are weaker for immigrants from richer countries. It is possible that in richer countries, compulsory schooling laws and better school quality help offset some of the disadvantages of arriving in the U.S. at a later age. However, the 2SLS estimate of the effect of one unit of English-speaking ability on wages is approximately the same for both richer and poorer countries – about 30 percent. Paralleling the OLS estimates, the return to English proficiency appears to be lower among immigrants from richer countries, but this gap is not significantly different from zero.

Table 7 restricts the analysis to the Caribbean region. Within this region, there are both English- and non-English-speaking countries. Restricting attention to this region yields a sample that is more similar in terms of geography, race, colonial history and GDP. Panel A for the entire Caribbean region suggests a 2SLS return of 44 percent for each unit of English-speaking ability. (Note that the standard errors are much higher now, due to the drastically reduced sample size.) Panels B to E do paired contrasts as an attempt to control better for GDP and similar returns to English-speaking ability are found.

2. Do parents factor in child's language-learning ability in the migration decision?

The second alternative hypothesis is that parents from non-English-speaking countries may factor their children's ages into the migration decision in a way that is different from parents from English-speaking countries. For example, the former may systematically enter when their children are younger because they realize the language-learning disadvantage their children would suffer if they do otherwise. Because of this, the distribution of parental characteristics across age at arrival may differ between English- and non-English-speaking countries. The 2SLS estimate may reflect not only the true effect of English-language proficiency, but also, the effects of differences in parental characteristics.

To assess this, we compare the age-at-arrival distribution of the two groups. Parents from non-English-speaking countries may factor their children's ages into the migration decision in a way that is different from parents from English-speaking countries. A component of the assumption that the immigrants from English-speaking countries serve as a good control is a certain similarity in the characteristics of the immigrants' parents (holding age at arrival fixed). However, immigrant parents optimizing the income of their "dynasty" should take into account the effect of language acquisition on earnings. In particular, parents coming from non-English-speaking countries should time their migration so that their children are younger when they arrive.[19] We might expect this to also affect, consequently, the distribution of parental characteristics across age at arrival.

Figure 5 shows the distribution of age at arrival for the treatment and control groups. Each point on the diamond-marker (square-marker) line gives the proportion of the immigrants from non-English-speaking countries (English-speaking countries) that arrived in the U.S. at that particular age. The two lines are similar. This suggests that although parents' migration decision may be sensitive to children's age, this sensitivity does not vary by English- and non-English-speaking country. It is not the case that parents from non-English speaking countries are *more* likely than parents from English-speaking countries to migrate when their children are very young, understanding that older children have a language-learning disadvantage. Had this been the case, there would have been more mass in the younger ages for the immigrants from non-English-speaking countries; Figure 5 shows that the reverse is true in our sample.

B. Contribution of Education to the Effect of English-Language Skills on Wages

In our sample, the causal effect of English-language proficiency on earnings is itself largely mediated by education, and is not due to a large direct effect of language on the marginal

[19] There is anecdotal evidence that many immigrants time their immigration before their fertility, but, as the anecdotes go, this has to do with the residency status of their children and not language acquisition.

product of labor. This is not a surprising conclusion given three important pieces of evidence: (1) we found a large and positive effect of English-language proficiency on education (see Section III.B.2); (2) a large literature has demonstrated substantial causal returns to education (see Card (1995) for a review); and (3) we found a large and positive effect of English-language proficiency on wages (see Section III.B.1).

The key question is to what extent (1) is generating (3). In this section, we address this issue by incorporating education directly into the wage regressions from above. We do this in two ways. First, we partial out the effect of schooling on wages using rates of return suggested by previous research. Second, we treat education as an exogenous control in 2SLS. Both approaches indicate that educational attainment is at the center of the observed language-wage relationship.

The resulting estimates are shown in Table 8. We start with the base specification for wages, shown in Column 1 (which summarizes Table 3); an additional point of English-proficiency brings about a 0.33 increase in log wages. In contrast, including education in the specification yields an estimate of the same effect that is lower by at least a factor of three. Using returns to schooling closer to those favored by our data, we find the estimated effect is lower by about a factor of ten. That is, approximately 90 percent of the effect of English-language skills on wages works through changing educational attainment. The remaining 10 percent may be due to other channels, such as the improved ability to communicate with customers and co-workers, although we cannot reject the hypothesis that *all* of the wage effect is mediated by schooling.

The large role that education plays in the effect of English-speaking ability on wages is not surprising given the changes in the mean and distribution of educational attainment that we found earlier. Better English-speaking ability induces immigrants who would otherwise complete eleven or fewer years of schooling to get at least their high school degree. College graduates earn more than high school graduates and dropouts, and this disparity has become more pronounced in recent decades.

C. Magnitude of the IV Estimate Compared to the OLS Estimate

One puzzle regarding our results is that IV estimate of the return to language skills is higher than the OLS estimate; the ability bias story in which omitted ability affects both earnings capacity and language acquisition predicts the reverse. In this subsection, we discuss two potential explanations: measurement error in the language skills measure and differences in the weighting function underlying the OLS and IV estimates.[20]

1. Is IV capturing individuals at a different part of the distribution than OLS?

First, the IV estimate uses only the variation in language skills that is correlated with the instrument whereas the OLS estimate uses all the variation. That is, IV puts more weight on individuals whose language skills are more affected by the instrument (Angrist and Imbens (1995)). In contrast, OLS weighs individuals in proportion to their contribution to the total change in language skills, irrespective of the instrument. To the extent that the marginal return to language skills for individuals more affected by the instrument differs systematically from those less affected, then the coefficient estimated using OLS will differ from that using IV.

Recall that there is no clear scaling *a priori* for our ordinal measure of language skills. It may be that the return to moving from speaking English "not at all" to speaking "not well" is different from the return from moving from "well" to "very well". Our estimates of the CDF differences arising from the binary instrument were presented in Table 2. The binary instrument

[20] These explanations have also been offered for why IV estimates of the returns to years of schooling are higher than their OLS counterparts. See Card (1995) for an overview.

shifted the CDF up (towards higher English-language proficiency) at every point in the distribution. However, most of the "mass" moved into the highest category, "speaks English very well", *i.e.*, the principal effect of arriving to the U.S. at a young age is to bring individuals who speak English well across the margin to very well. Thus, IV would yield a higher estimate than OLS if the greatest gains from language proficiency come from later steps towards proficiency. However, in our sample, OLS estimates of the marginal return at each point of English-speaking ability do not suggest nonlinearities in the returns to language skills.[21] Thus there is no direct support for the idea that the higher IV estimate is due to a simple reweighting of heterogeneous effects.

2. What is the extent of measurement error?

Second, there may be measurement error in the language skills measure. Let an individual's true, latent language skills be x^* and observed language skills be x, such that

(5) $$x = x^* + u$$

(subscript i has been suppressed). Suppose the true relationship between log wages (y) and language skills is

(6) $$y = \alpha + \beta x^* + \varepsilon$$

(for expositional ease, this is a bivariate form of Equation 3). Equation 6 satisfies the assumptions of the classical regression model. The researcher estimates the model using x instead of x^*. The resulting OLS estimate will be biased since the regressor is correlated with the error term:

---

[21] An OLS regression of wages on each point of English-speaking ability yields a coefficient of 0.1921 (standard error of 0.0524) for moving from no English to speaks English not well, 0.2651 (0.0264) for moving from not well to well, and 0.2046 (0.0153) for moving from well to very well. An F-test cannot reject the null hypothesis that the three coefficients are equal.

(7) $$b_{OLS} = \beta\left[\frac{Var(x^*) + Cov(x^*,u)}{Var(x^*) + Var(u) + 2Cov(x^*,u)}\right].$$

In the case of classical measurement error, with Cov(x*,u) = 0, we get the standard result of attenuation bias in the OLS estimate. The greater the noise, Var(u), the greater the bias towards zero. Thus, classical measurement error can explain why our IV estimate of the returns to language is higher than our OLS estimate. By instrumenting for the language measure (with *k*, an interaction between age at arrival and non-English-speaking country), we have likely purged some of the response noise from the language measure. This mitigates the attenuation bias, thus leading to higher IV estimates.

Nonclassical measurement error, with Cov(x*,u) ≠ 0, might also be a concern. When the latent explanatory variable is noisily measured in a few discrete categories, or has a lower or upper bound, in general both OLS and IV estimates will be inconsistent. Unfortunately, this is exactly the case with language measures based on Census questionnaires. The U.S. Census measures English-speaking ability in four discrete groups, whereas true language skills might more naturally be measured on a continuous scale. As well, data in the Census are self-reported. When Cov(x*,u) ≠ 0, the OLS estimate will biased as shown in Equation 7. Moreover, the IV estimate will be biased. Let *k* be an instrument for language skills, satisfying the criteria Cov(k,x*) ≠ 0 and Cov(k,ε) = 0. Write *k* as

(8) $$k = x^* + q$$

and let the error terms (ε, u and q) be uncorrelated. The IV estimate is just the indirect least squares estimate (*i.e.*, the ratio of the reduced-form effect on earnings and the reduced-form effect on language), and it can be shown that

(9) $$b_{IV} = \beta\left[\frac{Var(x^*) + Cov(x^*,q)}{Var(x^*) + Cov(x^*,q) + Cov(x^*,u)}\right].$$

How might a correlation between x* and u arise? Assume that the true language variable x* is continuous on the interval [0, 3]. Suppose respondents know their x*, but must categorize themselves as x = 0, 1, 2 or 3. If respondents are distributed uniformly, or are distributed symmetrically around the middle, then Cov(x*,u) > 0; intuitively, this is because there is more rounding up than rounding down (Berman, Lang and Siniver (2000)). Next, suppose that respondents sometimes misreport their x*. To the extent that there are many people at the bounds (0 or 3), then there will be a spurious negative relationship between x* and u: at the lower bound, measurement error will more likely be too positive (individuals have less room to under-report) and at the upper bound, it will more likely be too negative (individuals have less room to over-report).[22] Finally, suppose that in reality x* is continuous on the interval [0,4], but because of the Census' limited categories, individuals with language skill exceeding 3 are coded as 3. This "topcoding" will make Cov(x*,u) < 0, since at the upper bound individuals have less room to over-report. This might be of concern in our sample, where 83% of immigrants from non-English-speaking countries place themselves in the highest category, x = 3. It seems likely that within this category there are individuals with substantially better language skills than others.

If the lower and upper bounds induce a negative correlation between x* and u that exceeds the positive correlation induced by the rounding, then nonclassical measurement error can help explain why the IV estimate is higher than the OLS estimate – OLS is downward biased and IV is upward biased. Several methods have been proposed to correct for nonclassical measurement error, including using external validation data sets (*e.g.*, Card (1996)), restricting analysis to observations where two reports of the mismeasured variable agree (*e.g.*, Black,

[22] Naturally, if misreporting tends to occur only in particular parts of the language distribution or in a particular direction, then the sign of the bias on the IV estimate is ambiguous (for an example, see Kane *et al* (1999)).

Berger and Scott (2000)) and using general method of moments when two reports exist (*e.g.*, Kane, Rouse and Staiger (1999)). We adopt the first method; we emphasize that our intent is to get a rough idea of the extent of measurement error.

Our validating data set is the 1992 National Adult Literacy Survey (NALS).[23] The NALS was designed to study the nature and extent of literacy among adults in the U.S. (see National Center for Educational Statistics (1997)). Respondents answered background questions (including the Census language question verbatim) and took a 45-minute literacy test. The literacy test score is an appealing measure of English-language skills because it is based on an objective test (instead of a self-assessment), and also because it is measured in finer gradations (instead of four broad categories). To proceed, we construct the ordinal measure of language skills exactly as we did for the Census data based on the respondents' self-assessment of their own English-speaking ability – this is x. The mean is 2.4382, standard deviation is 0.8446 and the range is 0 to 3 (integer values only). We take the literacy test score to be the true measure of language skills – this is x*.[24] The mean is 2.5477, standard deviation is 0.6653 and the range is

[23] We do not use the NALS for all our analysis because of the paucity of observations. The NALS surveyed approximately 13,000 individuals, but less than 300 satisfy all the data restrictions described in Section II. The NALS data used below has 267 observations. They are immigrants from non-English-speaking countries who arrived to the U.S. between 1962 and 1981and are currently aged 23 to 38. We require non-missing literacy test score and self-assessment of English-speaking ability, but not non-missing wages.

[24] We can also let the test score measure be a noisy measure of true language skills. Under the assumption that the measurement errors in the self-assessment is and the test score are uncorrelated with each other and the error in the wage regression, then we can use methods described in Kane *et al.* (1999) and Black *et al.* (2000) to correct our estimates. If the measurement errors are correlated with each other or the wage regression, *e.g.*, if the two language variables encapsulate ability, then the following analysis using NALS data should be viewed as suggestive rather than definitive evidence on the role of measurement error. What is important is the test score appears to be a higher

0.7 to 3.9 (continuous values).[25]

We find that the signal-to-total variance ratio is 0.51. We can use this ratio to correct our estimates of the return to language for measurement error; we require that the relationship between x* and x in the NALS data applies to the Census data. Recall from Table 3 that the OLS estimate of the effect of language on earnings was 22%. Using Equation 7 to correct for measurement error, the OLS estimate doubles to 44%. We note that Cov(x*,u=x-x*) is -0.0808, Var(x*) is 0.4426 and Var(u) is 0.4323. The point estimate of the covariance between the true language measure and the measurement error is negative, although small in magnitude relative to the total noise.

When Cov(x*,u) $\neq$ 0, even the IV estimate will be inconsistent. Our IV estimate of the effect of language on earnings was 33% (from Table 3). Using Equation 8 to correct for measurement error, the IV estimate falls to 27%. This is lower than the corrected OLS estimate, 44%. This 17-percentage-point difference may be attributable to the fact that the OLS estimate does not correct for endogeneity while the IV estimate does. The upward bias of the OLS estimate is consistent with a significant role for the ability bias story. This upward bias is apparently masked by the severe downward bias associated with measurement error in the language variable based on the Census language question. Since many researchers studying the effects of language skills rely on data sets with the same survey instrument to measure language, this finding has widespread implications.[26] In particular, it would be difficult to make inferences

---

quality measure of language skills.

[25] We have divided test scores by 100. In theory, test scores can range from 0 to 500, but in our sample they took on a narrower set of values.

[26] The censuses of various other countries use the U.S. Census language question, including Australia, Canada and Israel. Additionally, the CPS in the U.S. also uses the Census language question.

about the effects of language skills without addressing both endogeneity and errors-in-variable (including nonclassical measurement error).

V. Conclusions

We find a significant positive effect of English-language skills on wages among individuals from the 1990 Census who immigrated to the U.S. as children. We control for non-language effects of age at arrival with immigrants from English-speaking countries. The estimated effect using our IV strategy is greater in magnitude than that suggested by regression strategies that do not address endogeneity and measurement error. We find evidence of substantial downward bias in the OLS estimate due to measurement error and somewhat smaller upward bias due to endogeneity.

Much of the effect of English-language skills appears to be mediated by years of schooling. Better English-language skills induce immigrants who would otherwise drop out with the equivalent of junior high or some high school education to at least complete their high school degree.

Our findings suggest that timing of migration and its effect on English-language skills are critical to a variety of important outcomes, and policymakers should be cognizant of this. Since much of the effect of English-language skills is through increased years of schooling, adult English-language classes may be insufficient to help these immigrants' wages to converge to those of natives. Instead, programs aimed at junior-high-school-aged and high-school-aged children may be more effective. Future work will explore in greater detail the policies and programs that may be most effective in mitigating the effect of poor English skills on the school-drop-out rates of immigrants.

References

Aigner, Dennis J., "Regression with a Binary Independent Variable Subject to Errors of Observation," *Journal of Econometrics* 1 (1973): 49-60.

Angrist, Joshua D. and Victor Lavy, "The Effect of a Change in Language of Instruction on the Returns to Schooling in Morocco," *Journal of Labor Economics* 9 (1991): 149-170.

Angrist, Joshua D. and Guido W. Imbens, "Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity," *Journal of the American Statistical Association* 90 (1995): 431-442.

Berman, Eli, Kevin Lang and Erez Siniver, "Language-Skill Complementarity: Returns to Immigrant Language Acquisition," mimeo, 2000.

Black, Dan, Mark Berger and Frank Scott, "Bounding Estimates with Non-Classical Measurement Error," *Journal of the American Statistical Association* 95 (2000): 739-748.

Borjas, George, "The Economic Analysis of Immigration," in O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics*, Amsterdam: Elsevier Science Publishing, 1999.

Borjas, George, "The Economics of Immigration," *Journal of Economic Literature* 32 (1994): 1661-1717.

Card, David "Earnings, Schooling, and Ability Revisited," in S. Polachek (ed.), *Research in Labor Economics*, Greenwich, CT: JAI Press, 1995.

Card, David "The Effect of Unions on the Structure of Wages: A Longitudinal Analysis," *Econometrica* 64 (1996): 957-979.

Chiswick, Barry R., "Speaking, Reading and Earnings among Low-Skilled Immigrants," *Journal of Labor Economics* 9 (1991): 149-170.

Chiswick, Barry R., "Hebrew Language Usage: Determinants and Effects on Earnings among Immigrants in Israel," *Journal of Population Economics* 11 (1998): 253-271.

Chiswick, Barry R. and Paul W. Miller, "Language in the Labor Market: The Immigrant Experience in Canada and the United States," in Barry R. Chiswick (ed.), *Immigration, Language and Ethnic Issues: Canada and the United States*, Washington, DC: American Enterprise Institute, 1992.

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

Chiswick, Barry R. and Paul W. Miller, "Language Skills and Earning among Legalized Aliens," *Journal of Population Economics* 12 (1999): 63-89.

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

Friedberg, Rachel M., "Immigration and the Labor Market," MIT Ph.D. Dissertation, 1993.

Friedberg, Rachel M., "You Can't Take it with You? Immigrant Assimilation and the Portability of Human Capital," *Journal of Labor Economics* 18 (2000): 221-251.

Kane, Thomas J., Cecilia E. Rouse and Douglas Staiger, "Estimating Returns to Schooling When Schooling is Misreported," NBER Working Paper No. 7235, 1999 (http://www.nber.org).

Katz, Lawrence F. and Kevin M. Murphy, "Changes in Relative Wages, 1963-1987: Supply and Demand Factors," *Quarterly Journal of Economics* 107 (1992): 35-78.

Kossoudji, Sherrie A., "English Language Ability and the Labor Market Opportunities of Hispanic and East Asian Immigrant Men," *Journal of Labor Economics* 6 (1988): 205-228.

Lee, Jong-Wha and Robert J. Barro, "Schooling Quality in a Cross Section of Countries," NBER Working Paper No. 6198, 1997 (http://www.nber.org).

McManus, Walter, William Gould and Finis Welch, "Earnings of Hispanic Men: The Role of English Language Proficiency," *Journal of Labor Economics* 2 (1983): 101-130.

National Center for Educational Statistics, *Technical Report and Data File Users Manual for the National Adult Literacy Survey*, Washington, D.C.: U.S. Department of Education NCES 97-060, 1997.

Newport, Elissa L., "Maturational Constraints on Language Learning," *Cognitive Science*, 14 (1990): 11-28.

Ruggles, Steven, Matthew Sobek et al., *Public Use Microdata Series: Version 2.0*, Minneapolis: Historical Census Projects, University of Minnesota, 1997 (http://www.ipums.umn.edu).

Ruiz-de-Velasco, Jorge, Michael Fix and Beatriz Chu Clewell, *Overlooked and Underserved: Immigrant Students in U.S. Secondary Schools*, Washington, DC: Urban Institute, 2000.

Schultz, T. Paul, "Immigrant Quality and Assimilation: A Review of the U.S. Literature," *Journal of Population Economics*, 11 (1998): 239-252.

Secada, Walter G. *et al.*, *No More Excuses: The Final Report of the Hispanic Dropout Project*, Washington, DC: Department of Education, 1998.

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

Tanier, Evelina, "English Language Proficiency and Earnings among Foreign-born Men," *Journal of Human Resources* 23 (1988): 108-122.

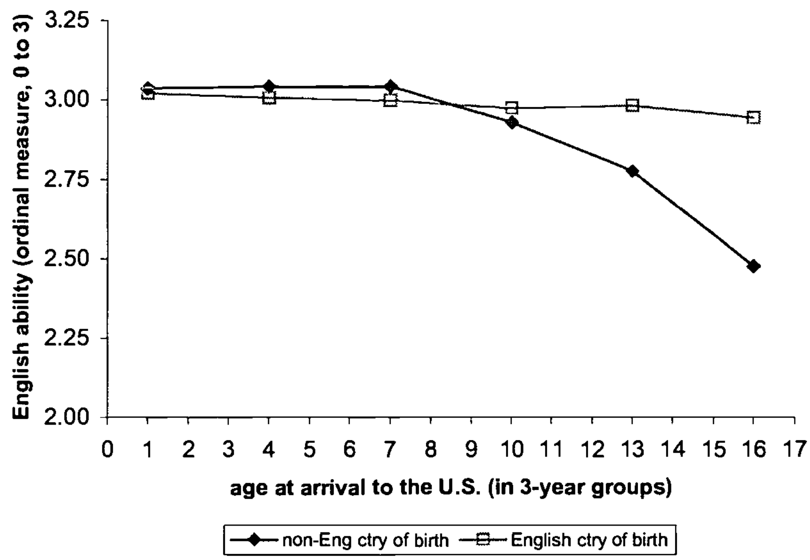
U.S. Department of Commerce, Bureau of the Census, *Social Characteristics: 1990* (http://www.census.gov).

U.S. Department of Commerce, Bureau of the Census, *Profile of Selected Social Characteristics: 2000* (http://www.census.gov).

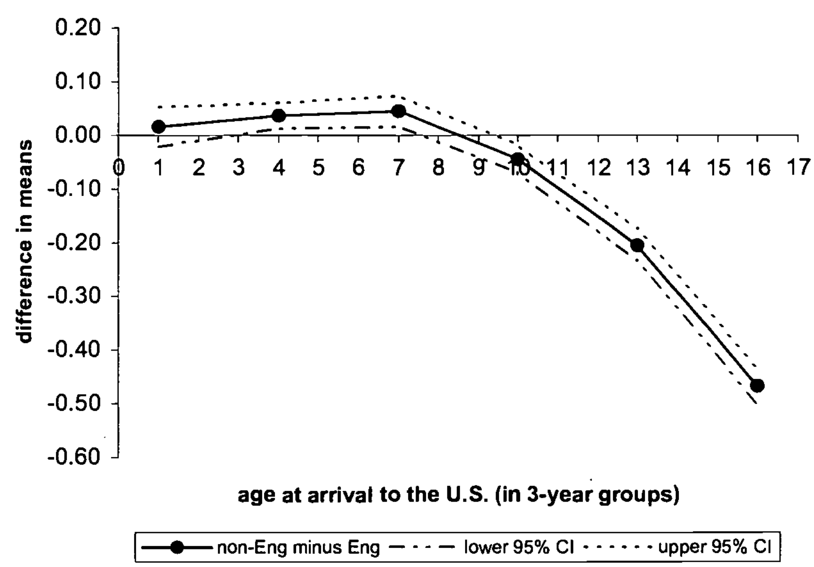
World Almanac and Book of Facts, 1999, New York, NY: World Almanac Books, 1999.

Figure 1. English-Speaking Ability by Age at Arrival

Panel A. Regression-Adjusted Means



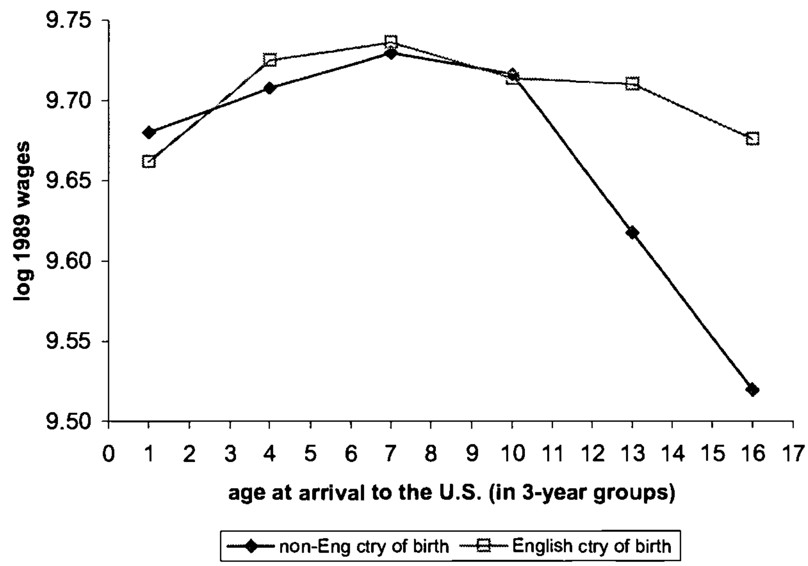
Panel B. Difference in Means



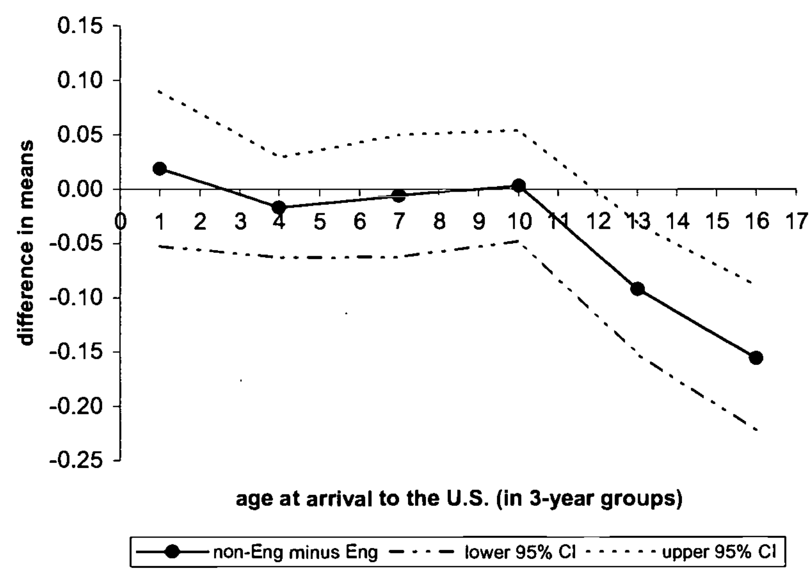
Notes: Data from 1990 IPUMS. Sample size is 66,584 (comprised of individuals who arrived to the U.S. by age 17 between 1960 and 1974 and currently aged 25 to 38). English ordinal measure: 0 = no English, 1 = not well, 2 = well and 3 = very well. Means have been regression-adjusted for age, race/ethnicity and female dummies.

Figure 2. Log Annual Wages by Age at Arrival

Panel A. Regression-Adjusted Means



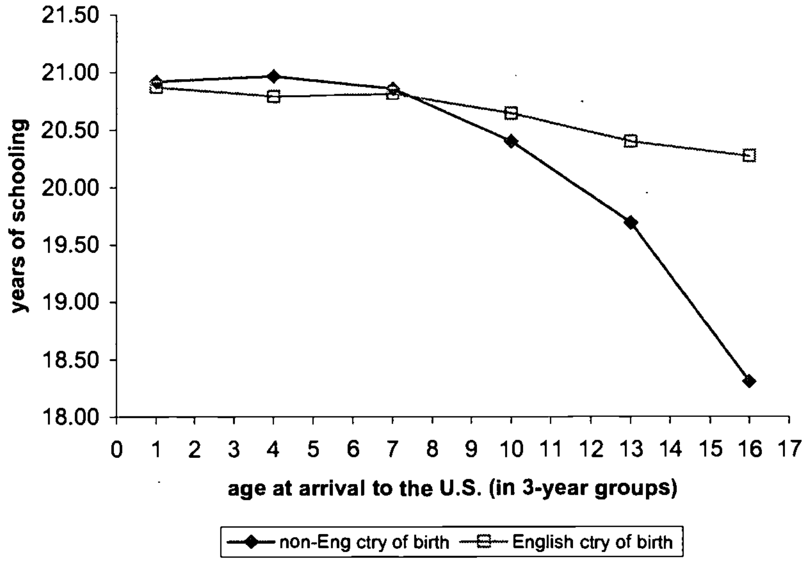
Panel B. Difference in Means



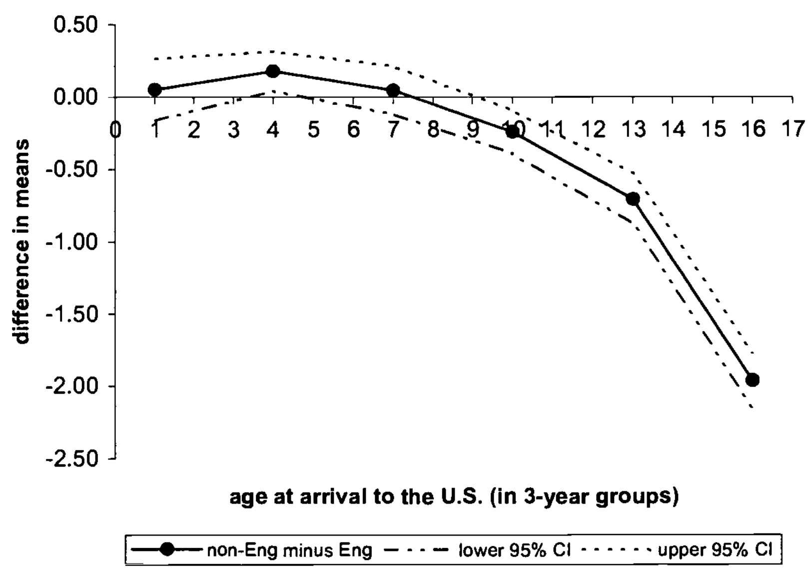
Notes: Data from 1990 IPUMS. Sample size is 47,422 (comprised of individuals who arrived to the U.S. by age 17 between 1960 and 1974 and currently aged 25 to 38). Means have been regression-adjusted for age, race/ethnicity and female dummies.

Figure 3. Years of Schooling by Age at Arrival

Panel A. Regression-Adjusted Means



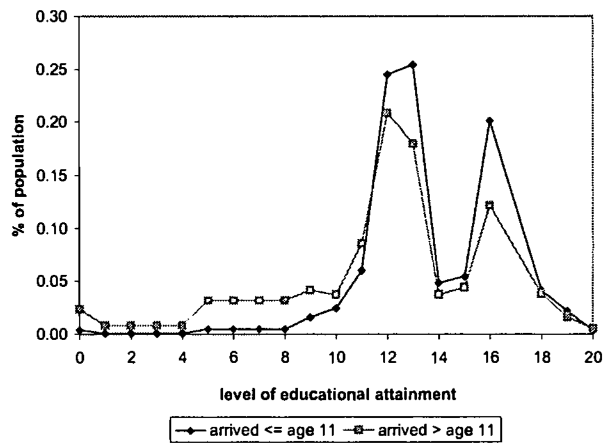
Panel B. Difference in Means



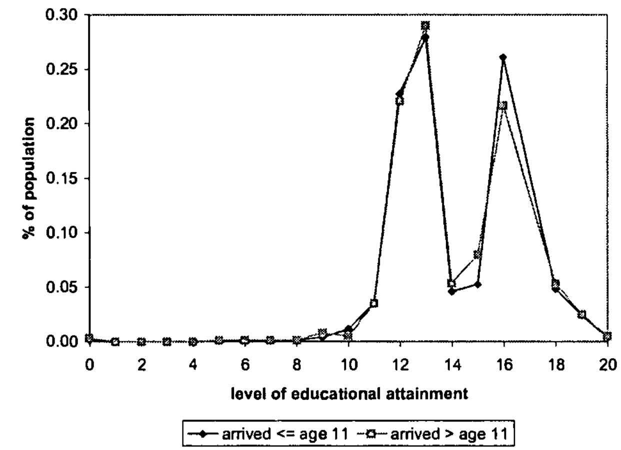
Notes: Data from 1990 IPUMS. Sample size is 65,214 (comprised of individuals who arrived to the U.S. by age 17 between 1960 and 1974 and currently aged 25 to 38). Means have been regression-adjusted for age, race/ethnicity and female dummies.

Figure 4. Probability Distribution Function of Educational Attainment

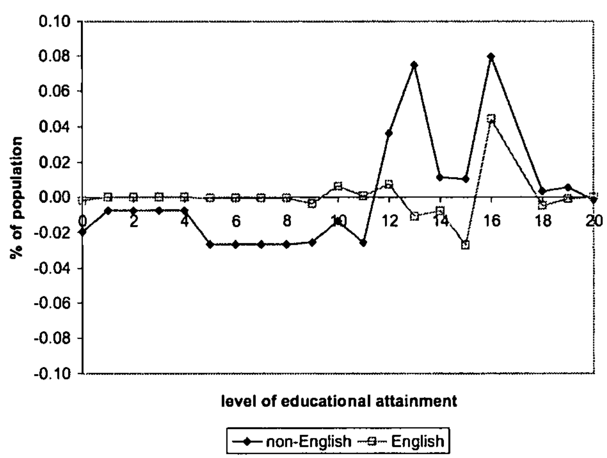
Panel A. PDF for non-English-speaking countries



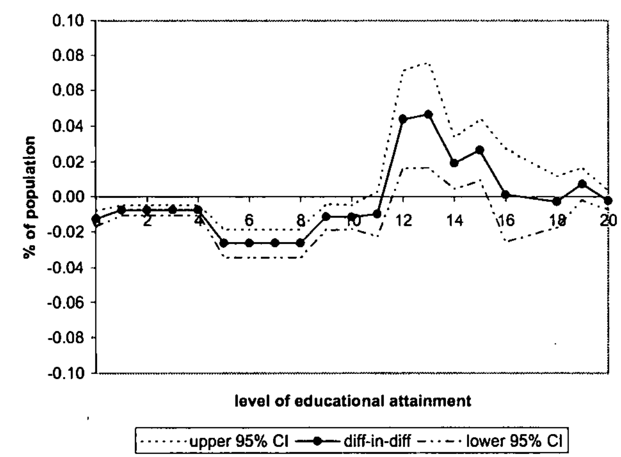
Panel B. PDF for English-speaking countries



Panel C. Difference in PDF between Young and Old

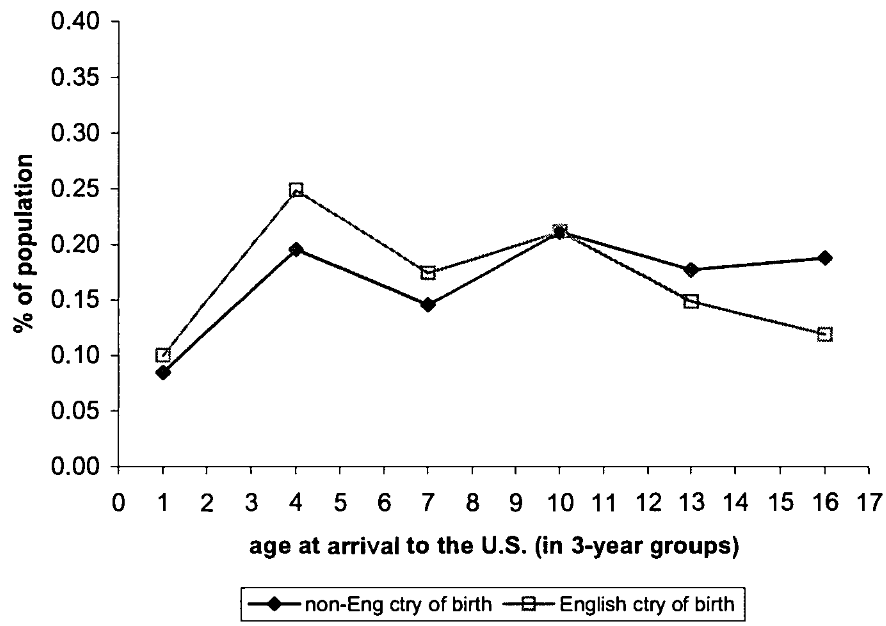


Panel D. Difference-in-Differences in PDF, Regression-Adjusted



Notes: Panels A, B and C display the unconditional PDFs. Panel D displays a double difference PDF that has been regression-adjusted for age, race/ethnicity and female dummies. The level of educational attainment coding used above is based on the 1990 Census coding: 0 = less than first grade; 1, 2, 3 and 4 = first through fourth grade (probabilities rescaled to a "per grade" basis for comparability); 5, 6, 7 and 8 = fifth through eighth grade (rescaled for comparability); 9 = ninth grade; 10 = tenth grade; 11 = eleventh grade and twelfth grade without diploma; 12 = high school diploma; 13 = some college (no degree); 14 = associate degree (occupational program); 15 = associate degree (academic program); 16 = Bachelor's degree; 18 = Master's degree; 19 = professional school degree; and 20 = doctorate degree.

Figure 5. Probability Distribution Function of Age at Arrival



Notes: Data from 1990 IPUMS. Sample size is 66,584 (comprised of individuals who arrived to the U.S. by age 17 between 1960 and 1974 and currently aged 25 to 38), of which 57,106 are from a non-English-speaking country of birth and the remaining 9,478 are from an English-speaking country of birth.

Table 1. Descriptive Statistics

	immig from non-English-spking ctries			immig from English-spking ctries		
	overall	arrived aged 0-11	arrived aged 12-17	overall	arrived aged 0-11	arrived aged 12-17
	(1)	(2)	(3)	(4)	(5)	(6)
log annual wages	9.6699	9.6723	9.6652	9.7648	9.7363	9.8426
	(0.9449)	(0.9424)	(0.9499)	(0.9537)	(0.9573)	(0.9397)
English-speaking ability variables						
ordinal measure (scale	2.7693	2.8928	2.5259	2.9863	2.9858	2.9878
of 0 to 3, 3=best)	(0.5545)	(0.3746)	(0.7397)	(0.1323)	(0.1383)	(0.1143)
speaks English	0.0083	0.0024	0.0200	0.0000	0.0000	0.0000
not at all (0)	(0.0909)	(0.0491)	(0.1400)	(0.0000)	(0.0000)	(0.0000)
speaks English	0.0399	0.0151	0.0889	0.0020	0.0026	0.0005
not well (1)	(0.1958)	(0.1219)	(0.2846)	(0.0448)	(0.0507)	(0.0222)
speaks English	0.1258	0.0698	0.2363	0.0096	0.0090	0.0112
well (2)	(0.3317)	(0.2548)	(0.4248)	(0.0977)	(0.0947)	(0.1054)
speaks English	0.8259	0.9127	0.6548	0.9884	0.9884	0.9883
very well (3)	(0.3792)	(0.2822)	(0.4754)	(0.1073)	(0.1072)	(0.1077)
control variables						
age at arrival	8.9789	6.1663	14.5168	8.2438	6.0229	14.3058
	(4.8341)	(3.1853)	(1.7770)	(4.6251)	(3.1179)	(1.7415)
age	30.4483	29.1236	33.0567	30.1490	29.1121	32.9793
	(3.6630)	(3.1822)	(3.1048)	(3.5596)	(3.1151)	(3.1408)
white	0.8893	0.8927	0.8825	0.7243	0.8163	0.4732
	(0.3138)	(0.3095)	(0.3220)	(0.4469)	(0.3873)	(0.4994)
black	0.0425	0.0429	0.0418	0.2478	0.1603	0.4864
	(0.2017)	(0.2025)	(0.2002)	(0.4317)	(0.3670)	(0.4999)
Asian/other non-white	0.0682	0.0644	0.0757	0.0279	0.0234	0.0405
race	(0.2521)	(0.2455)	(0.2645)	(0.1648)	(0.1511)	(0.1971)
Hispanic	0.5394	0.4744	0.6674	0.0170	0.0149	0.0227
	(0.4985)	(0.4994)	(0.4711)	(0.1293)	(0.1213)	(0.1489)
female	0.4559	0.4657	0.4367	0.4937	0.4801	0.5309
	(0.4981)	(0.4988)	(0.4960)	(0.5000)	(0.4997)	(0.4992)
schooling variables						
years of schooling	13.0773	13.6567	11.9282	14.2124	14.2324	14.1576
	(3.2525)	(2.6293)	(3.9828)	(2.2605)	(2.2370)	(2.3233)
completed high school	0.7979	0.8718	0.6514	0.9432	0.9433	0.9430
	(0.4016)	(0.3343)	(0.4765)	(0.2314)	(0.2313)	(0.2319)
completed college	0.2391	0.2684	0.1812	0.3276	0.3380	0.2991
	(0.4266)	(0.4431)	(0.3852)	(0.4694)	(0.4731)	(0.4580)
Number of observations	40,258	26,490	13,768	7,164	5,309	1,855
N for schooling variables	39,647	26,154	13,493	7,097	5,260	1,837

Notes: Means weighted by IPUMS weights. Sample is as follows: 1990 IPUMS, arrived to the U.S. by age 17 between 1960 and 1974, is currently aged 25 to 38 and with nonmissing language and wage variables.

Table 2. Difference-in-Differences with Binary Treatment Variable

	speaks English not well, well or very well *mean for old & non-Eng = 0.9800*		speaks English well or very well *mean = 0.8911*		speaks English very well *mean = 0.6548*		English ability (ordinal measure, 0 to 3) *mean = 2.5259*		log annual wages *mean = 9.6652*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
arrived young*non-English-speaking country of birth	**0.0176** ***	**0.0142** ***	**0.0935** ***	**0.0794** ***	**0.2578** ***	**0.2188** ***	**0.3689** ***	**0.3124** ***	**0.1134** ***	**0.1221** ***
	(0.0013)	**(0.0012)**	**(0.0031)**	**(0.0031)**	**(0.0057)**	**(0.0060)**	**(0.0080)**	**(0.0082)**	**(0.0298)**	**(0.0302)**
arrived young (aged 0 to 11)	0.0000	0.0018 ***	-0.0021 **	0.0007	0.0001	0.0101 **	-0.0020	0.0125 **	-0.1063 ***	0.0206
	(0.0000)	(0.0006)	(0.0009)	(0.0021)	(0.0030)	(0.0045)	(0.0034)	(0.0059)	(0.0277)	(0.0295)
non-English-speaking country of birth	-0.0200 ***	-0.0109 ***	-0.1084 ***	-0.0681 ***	-0.3334 ***	-0.2178 ***	-0.4619 ***	-0.2968 ***	-0.1774 ***	-0.1277 ***
	(0.0013)	(0.0009)	(0.0029)	(0.0027)	(0.0051)	(0.0058)	(0.0073)	(0.0075)	(0.0252)	(0.0271)
Age dummies	N	Y	N	Y	N	Y	N	Y	N	Y
Race, ethnicity & sex dummies	N	Y	N	Y	N	Y	N	Y	N	Y
Adjusted R-squared	0.0097	0.0149	0.0472	0.0680	0.1267	0.1851	0.1179	0.1618	0.0017	0.0796
Number of observations	47,422	47,422	47,422	47,422	47,422	47,422	47,422	47,422	47,422	47,422

Notes: Weighted by IPUMS weights. Robust standard errors in parentheses. Single asterisk denotes statistical significance at the 90% level of confidence, double 95%, triple 99%.
Sample is as follows: 1990 IPUMS, arrived to the U.S. by age 17 between 1960 and 1974, is currently aged 25 to 38 and with nonmissing language and wage variables.
English-speaking ability ordinal measure: 0 = no English, 1 = not well, 2 = well and 3 = very well.

Table 3. Effect on Log Annual Wages -- Base Results

	English Ability (ordinal measure, 0 to 3) *mean for non-Eng ctry = 2.7693*		Log Annual Wages *mean for non-English-speaking countries = 9.6699*					
	OLS 1st stage (1)	OLS 1st stage (2)	OLS reduced-form (3)	OLS reduced-form (4)	OLS (5)	OLS (6)	2SLS 2nd stage (7)	2SLS 2nd stage (8)
endogenous regressor								
English-speaking ability					0.2225 ***	0.2219 ***	0.3286 ***	0.3335 ***
(scale of 0 to 3, 3=best)					(0.0093)	(0.0093)	(0.1060)	(0.1054)
excluded instrument in 2SLS 2nd stage								
max (0, age at arrival - 11) * non-	-0.0771 ***	-0.0776 ***	-0.0253 ***	-0.0259 ***				
English-speaking country of birth	(0.0021)	(0.0021)	(0.0082)	(0.0082)				
other exogenous regressors								
max (0, age at arrival - 11)	-0.0022 ***		-0.0054		-0.0121 ***		-0.0047	
	(0.0014)		(0.0080)		(0.0030)		(0.0082)	
Age at arrival dummies	Y	Y	Y	Y	Y	Y	Y	Y
Country of birth dummies	Y	Y	Y	Y	Y	Y	Y	Y
Age dummies	Y	Y	Y	Y	Y	Y	Y	Y
Race, ethnicity & sex dummies	Y	Y	Y	Y	Y	Y	Y	Y
Adjusted R-squared	0.2328	0.2352	0.1013	0.1014	0.1125	0.1125		
Number of observations	47,422	47,422	47,422	47,422	47,422	47,422	47,422	47,422

Notes: Weighted by IPUMS weights. Robust standard errors in parentheses. Single asterisk denotes statistical significance at the 90% level of confidence, double 95%, triple 99%.
Sample is as follows: 1990 IPUMS, arrived to the U.S. by age 17 between 1960 and 1974, is currently aged 25 to 38 and with nonmissing language and wage variables.
English-speaking ability ordinal measure: 0 = no English, 1 = not well, 2 = well and 3 = very well. Age at arrival has been demeaned to facilitate interpretation of the main effects.

Table 4. Effect on Log Annual Wages and Years of Schooling, Including and Excluding Mexico and Canada

	outcome = Log Annual Wages endogenous regressor = Language Ability			outcome = Years of Schooling endogenous regressor = Language Ability		
	1st stage	OLS	2SLS	1st stage	OLS	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. All Countries (Base)						
max (0, age at arrival - 11) * non-English-speaking country of birth	-0.0771 ***			-0.0841 ***		
	(0.0021)			(0.0019)		
English-speaking ability		0.2225 ***	0.3286 ***		1.9920 ***	4.0089 ***
		(0.0093)	(0.1060)		(0.0295)	(0.2279)
N	47,422	47,422	47,422	65,214	65,214	65,214
Panel B. Excluding immigrants from both Mexico and Canada						
max (0, age at arrival - 11) * non-English-speaking country of birth	-0.0443 ***			-0.0532 ***		
	(0.0022)			(0.0020)		
English-speaking ability		0.1847 ***	0.3428		1.8104 ***	3.5367 ***
		(0.0155)	(0.2290)		(0.0448)	(0.3995)
N	34,291	34,291	34,291	46,875	46,875	46,875
Panel C. Excluding immigrants from Mexico only						
max (0, age at arrival - 11) * non-English-speaking country of birth	-0.0434 ***			-0.0525 ***		
	(0.0021)			(0.0020)		
English-speaking ability		0.1840 ***	0.3499 *		1.8005 ***	3.3289 ***
		(0.0153)	(0.1940)		(0.0444)	(0.3707)
N	37,146	37,146	37,146	50,601	50,601	50,601
Panel D. Excluding immigrants from Canada only						
max (0, age at arrival - 11) * non-English-speaking country of birth	-0.0780 ***			-0.0844 ***		
	(0.0022)			(0.0019)		
English-speaking ability		0.2220 ***	0.3285 ***		1.9940 ***	4.1444 ***
		(0.0094)	(0.1274)		(0.0296)	(0.2489)
N	44,567	44,567	44,567	61,488	61,488	61,488

Notes: Weighted by IPUMS weights. Robust standard errors in parentheses. Single asterisk denotes statistical significance at the 90% level of confidence, double 95%, triple 99%. Sample is as follows: 1990 IPUMS, arrived to the U.S. by age 17 between 1960 and 1974, is currently aged 25 to 38 and no missing data for wages, English-speaking ability and GDP. All specifications include age at arrival main effect, and country of birth, age, race/ethnicity and sex dummies.

Table 5. Effect on Log Annual Wages -- School Quality Controls

	first stage (1)	reduced-form (2)	OLS (3)	2SLS 2nd stage (4)	N (5)
Panel A. Base (from Table 3)					
English-speaking ability (scale of 0 to 3, 3=best)			0.2219 *** (0.0093)	0.3335 *** (0.1054)	47,422
max (0, age at arrival - 11) * non-English-speaking country of birth	-0.0776 *** (0.0021)	-0.0259 *** (0.0082)			
Panel B. Control for GDP in Country of Birth					
English-speaking ability (scale of 0 to 3, 3=best)			0.2208 *** (0.0097)	0.3317 *** (0.0986)	40,552
max (0, age at arrival - 11) * non-English-speaking country of birth	-0.0908 *** (0.0029)	-0.0301 *** (0.0090)			
max (0, age at arrival - 11) * ln(per capita PPP GDP)	-0.0146 *** (0.0025)	-0.0018 (0.0050)	0.0032 (0.0046)	0.0031 (0.0046)	
Panel C. Control for School Expenditures in Country of Birth					
English-speaking ability (scale of 0 to 3, 3=best)			0.2173 *** (0.0101)	0.3628 ** (0.1755)	36,272
max (0, age at arrival - 11) * non-English-speaking country of birth	-0.0543 *** (0.0026)	-0.0197 ** (0.0095)			
max (0, age at arrival - 11) * ln(school exp per child)	0.0362 *** (0.0020)	0.0128 *** (0.0040)	0.0064 * (0.0036)	-0.0004 (0.0088)	
Panel D. Control for Teacher-Pupil Ratio in Country of Birth					
English-speaking ability (scale of 0 to 3, 3=best)			0.2174 *** (0.0100)	0.4031 *** (0.1344)	38,563
max (0, age at arrival - 11) * non-English-speaking country of birth	-0.0647 *** (0.0024)	-0.0261 *** (0.0087)			
max (0, age at arrival - 11) * ln(teacher-pupil ratio)	0.1094 *** (0.0053)	0.0256 *** (0.0098)	0.0046 (0.0095)	-0.0185 (0.0200)	
Panel E. Control for All Three "School Quality" Measures in Country of Birth					
English-speaking ability (scale of 0 to 3, 3=best)			0.2170 *** (0.0101)	0.3095 ** (0.1410)	36,272
max (0, age at arrival - 11) * non-English-speaking country of birth	-0.0674 *** (0.0030)	-0.0209 ** (0.0095)			

Notes: Weighted by IPUMS weights. Robust standard errors in parentheses. Single asterisk denotes statistical significance at the 90% level of confidence, double 95%, triple 99%. Sample is as follows: 1990 IPUMS, arrived to the U.S. by age 17 between 1960 and 1974, is currently aged 25 to 38 and no missing data for wages, English-speaking ability and the relevant "school quality" measure. Age at arrival and each "school quality" measure have been demeaned to facilitate interpretation of the main effects. All specifications include age at arrival, country of birth, age, race/ethnicity and sex dummies.

Table 6. Effect on Log Annual Wages -- High and Low GDP Countries

	first stage (1)	reduced-form (2)	OLS (3)	2SLS 2nd stage (4)	N (5)
Panel A. Base Case: All Countries with Nonmissing 1965 GDP Data					
English-speaking ability (scale of 0 to 3, 3=best)			0.2208 *** (0.0097)	0.3514 *** (0.1010)	40,552
max (0, age at arrival - 11) * non-English-speaking country of birth	-0.0830 *** (0.0023)	-0.0292 *** (0.0084)			
Panel B. Interactions of Key Regressors with High-GDP Country of Birth					
English-speaking ability			0.2329 *** (0.0105)	0.3375 *** (0.1248)	40,552
English-speaking ability * I(Above-median-GDP country of birth)			-0.0874 *** (0.0281)	-0.0690 (0.1995)	
max (0, age at arrival - 11) * non-English-speaking country of birth	-0.0822 *** (0.0035)	-0.0275 *** (0.0102)			
Z * I(Above-median-GDP country of birth)	0.0315 *** (0.0055)	0.0134 (0.0097)			
Panel C. Interactions of All Regressors with High-GDP Country of Birth					
English-speaking ability			0.2326 *** (0.0105)	0.3369 *** (0.1230)	40,552
English-speaking ability * I(Above-median-GDP country of birth)			-0.0872 *** (0.0281)	-0.0669 (0.2010)	
max (0, age at arrival - 11) * non-English-speaking country of birth	-0.0834 *** (0.0035)	-0.0279 *** (0.0102)			
Z * I(Above-median-GDP country of birth)	0.0338 *** (0.0054)	0.0141 (0.0097)			

Notes: Weighted by IPUMS weights. Robust standard errors in parentheses. Single asterisk denotes statistical significance at the 90% level of confidence, double 95%, triple 99%. Sample is as follows: 1990 IPUMS, arrived to the U.S. by age 17 between 1960 and 1974, is currently aged 25 to 38 and no missing data for wages, English-speaking ability and GDP. All specifications include age at arrival, country of birth, age, race/ethnicity and sex dummies. Panel **B** reports a specification with interactions of the endogenous regressor, the excluded instrument, and age at arrival effects with a dummy equal to one if the country of origin had above-median GDP in 1965. Panel C allows for interactions of all RHS variables with the above-median-GDP dummy.

Table 7. Effect on Log Annual Wages -- Caribbean Countries Only

	first stage (1)	reduced-form (2)	OLS (3)	2SLS 2nd stage (4)	N (5)
Panel A. All Caribbean					
English-speaking ability			0.2204 ***	0.4393 *	9,953
(scale of 0 to 3, 3=best)			(0.0258)	(0.2350)	
max (0, age at arrival - 11) * non-	-0.0638 ***	-0.0280 *			
English speaking country of birth	(0.0041)	(0.0150)			
Panel B. Jamaica (1965 PPP GDP = $2104) vs Puerto Rico (1965 PPP GDP = $4414)					
English-speaking ability			0.2257 ***	0.1859	3,165
(scale of 0 to 3, 3=best)			(0.0424)	(0.2522)	
max (0, age at arrival - 11) * non-	-0.0883 ***	-0.0164			
English speaking country of birth	(0.0081)	(0.0224)			
Panel C. Jamaica (1965 PPP GDP = $2104) vs Dominican Republic (1965 PPP GDP = $1271)					
English-speaking ability			0.1998 ***	0.3392	1,470
(scale of 0 to 3, 3=best)			(0.0550)	(0.3813)	
max (0, age at arrival - 11) * non-	-0.0631 ***	-0.0214			
English speaking country of birth	(0.0122)	(0.0241)			
Panel D. Jamaica (1965 PPP GDP = $2104) vs Cuba (1965 PPP GDP N/A)					
English-speaking ability			0.2270 ***	0.6240	5,745
(scale of 0 to 3, 3=best)			(0.0389)	(0.3857)	
max (0, age at arrival - 11) * non-	-0.0541 ***	-0.0338			
English speaking country of birth	(0.0051)	(0.0209)			
Panel E. Trinidad and Tobago (1965 PPP GDP = $6428) vs Puerto Rico (1965 PPP GDP = $4414)					
English-speaking ability			0.2354 ***	0.7026 **	2,753
(scale of 0 to 3, 3=best)			(0.0429)	(0.3295)	
max (0, age at arrival - 11) * non-	-0.0844 ***	-0.0593 **			
English speaking country of birth	(0.0086)	(0.0273)			

Notes: Weighted by IPUMS weights. Robust standard errors in parentheses. Single asterisk denotes statistical significance at the 90% level of confidence, double 95%, triple 99%. Sample is as follows: 1990 IPUMS, arrived to the U.S. by age 17 between 1960 and 1974, is currently aged 25 to 38 and no missing data for wages and English-speaking ability. All specifications include age at arrival, country of birth, age, race/ethnicity and sex dummies.

Table 8. Effect of English-Speaking Ability on Log Annual Wages, As Mediated by Years of Schooling

	Base	Various Assumptions about Returns to Schooling				Returns Estimated	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: 2SLS Estimates							
endogenous regressor							
English-speaking ability	0.3335 ***	0.1018	0.0622	0.0226	-0.0169	0.0183	0.0309
	(0.1054)	(0.1031)	(0.1028)	(0.1026)	(0.1024)	(0.1312)	(0.2511)
controls							
Years of Schooling	---	0.060 #	0.070 #	0.080 #	0.090 #	0.0811 ***	
						(0.0075)	
Dummies for Educational Attainment							Yes
Potential Years of Schooling in Origin Country							-0.0058
							(0.0113)
Potential Origin-Country Schooling * Indicators of School Quality							Yes
Panel B: OLS Estimates Using Same Sets of Controls							
English-speaking ability on Wages	0.2219 ***	0.1005 ***	0.0805 ***	0.0605 ***	0.0404 ***	0.0633 ***	0.0950 ***
	(0.0093)	(0.0093)	(0.0093)	(0.0094)	(0.0094)	(0.0100)	(0.0114)
Reduced form: k on Wages	-0.0259 ***	-0.0079	-0.0048	-0.0018	0.0013	-0.0011	-0.0011
	(0.0082)	(0.0080)	(0.0080)	(0.0079)	(0.0079)	(0.0079)	(0.0090)
First Stage: k on English-speaking ability	-0.0776 ***	-0.0776 ***	-0.0776 ***	-0.0776 ***	-0.0776 ***	-0.0605 ***	-0.0358 ***
	(0.0021)	(0.0021)	(0.0021)	(0.0021)	(0.0021)	(0.0022)	(0.0027)

Notes: Weighted by IPUMS weights. Robust standard errors in parentheses. Single asterisk denotes statistical significance at the 90% level of confidence, double 95%, triple 99%. Sample is as follows: 1990 IPUMS, arrived to the U.S. by age 17 between 1960 and 1974, is currently aged 25 to 38 and no missing data for wages, English-speaking ability and GDP. All specifications include age at arrival, country of birth, age, race/ethnicity and sex dummies. "k" denotes the excluded instrument. "#" indicates that the coefficient is constrained to reported value.

Appendix Table 1. Immigrants by Country of Birth

Panel A. Countries with English as an official language

1. English-speaking countries (=Control Group)

Rank by N	country	N	share of total N
1	Canada	3,775	39.8%
2	UK - England	2,242	23.7%
3	Jamaica	1,040	11.0%
4	Trinidad and Tobago	451	4.8%
5	UK - not specified	312	3.3%
6	UK - Scotland	298	3.1%
7	Guyana/British Guyana	180	1.9%
8	Australia	167	1.8%
9	Ireland	163	1.7%
10	Bermuda	162	1.7%
11	Barbados	113	1.2%
12	Belize/British Honduras	94	1.0%
13	South Africa	70	0.7%
14	Bahamas	60	0.6%
15	U.S. Virgin Islands	54	0.6%
16	New Zealand	48	0.5%
17	Grenada	37	0.4%
18	Northern Ireland	36	0.4%
19	St. Kitts-Nevis	32	0.3%
20	Liberia	31	0.3%
21	Antigua-Barbuda	30	0.3%
22	St. Vincent	18	0.2%
23	UK - Wales	16	0.2%
24	St. Lucia	16	0.2%
25	Zimbabwe	11	0.1%
26	British West Indies - n.s.	7	0.1%
27	Cayman Islands	6	0.1%
28	British Virgin Islands	5	0.1%
29	UK - Jersey	2	0.0%
30	Anguilla	2	0.0%
	Total English-spking obs	*9,478*	*100.0%*
	As % of total obs		*13.2%*

2. Other English-official countries (excluded from main analysis)

Rank by N	country	N	share of total N
1	Philippines	2,474	48.3%
2	Hong Kong	1,260	24.6%
3	India	493	9.6%
4	Guam	263	5.1%
5	Pakistan	102	2.0%
6	US territory - not specified	99	1.9%
7	Nigeria	63	1.2%
8	American Samoa	59	1.2%
9	Fiji	50	1.0%
10	Dominica	42	0.8%
11	Kenya	31	0.6%
12	Malta	31	0.6%
13	Tonga	25	0.5%
14	Singapore	23	0.4%
15	Uganda	23	0.4%
16	Tanzania	17	0.3%
17	Ghana	15	0.3%
18	Papua New Guinea	13	0.3%
19	Micronesia	9	0.2%
20	Zambia	6	0.1%
21	Marshall Islands	6	0.1%
22	Sierra Leone	6	0.1%
23	Mauritius	3	0.1%
24	Palau	3	0.1%
25	Gambia	2	0.0%
26	Gibraltar	2	0.0%
27	Seychelles	2	0.0%
28	Senegal	2	0.0%
29	Kiribati	1	0.0%
	Total other Eng-official obs	*5,125*	*100.0%*
	As % of total obs		*7.1%*

Notes: Table 1 continued on next page.

Appendix Table 1. Immigrants by Country of Birth (continued)

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

Rank by N	country	N	share of total N	Rank by N	country	N	share of total N
1	Mexico	15,035	26.3%	31	Israel/Palestine	278	0.5%
2	Cuba	6,554	11.5%	32	Costa Rica	263	0.5%
3	Germany excl. West G.	4,660	8.2%	33	Honduras	259	0.5%
4	Puerto Rico	4,404	7.7%	34	Iran	257	0.5%
5	Italy	2,358	4.1%	35	Turkey	244	0.4%
6	Japan	2,044	3.6%	36	South Korea	199	0.3%
7	West Germany	1,915	3.4%	37	Egypt	192	0.3%
8	Portugal	1,398	2.4%	38	Nicaragua	183	0.3%
9	France	1,265	2.2%	39	Lebanon	182	0.3%
10	Dominican Republic	1,238	2.2%	40	Vietnam	173	0.3%
11	Outside US, not specified	1,059	1.9%	41	Thailand	158	0.3%
12	Colombia	1,027	1.8%	42	Chile	153	0.3%
13	Korea excl. South Korea	962	1.7%	43	Indonesia	133	0.2%
14	Greece	832	1.5%	44	Morocco	130	0.2%
15	Poland	798	1.4%	45	Iraq	117	0.2%
16	Ecuador	588	1.0%	46	Africa - not specified	113	0.2%
17	Haiti	542	0.9%	47	Jordan	109	0.2%
18	Yugoslavia	515	0.9%	48	Czechoslovakia	107	0.2%
19	Spain	508	0.9%	49	Libya	105	0.2%
20	Taiwan	493	0.9%	50	Switzerland	97	0.2%
21	Argentina	484	0.8%	51	Romania	90	0.2%
22	El Salvador	434	0.8%	52	Austria	90	0.2%
23	Panama	412	0.7%	53	Bolivia	84	0.1%
24	China	390	0.7%	54	Russia	81	0.1%
25	Brazil	343	0.6%	55	Hungary	80	0.1%
26	Netherlands	337	0.6%	56	Uruguay	79	0.1%
27	Guatemala	326	0.6%	57	Subtotal, top 60 countries	55,765	97.7%
28	Peru	322	0.6%	58	Subtotal, other (87) countries	1,341	2.3%
29	Azores	285	0.5%	59	*Total non-Eng-spking obs*	*57,106*	*100.0%*
30	Venzuela	281	0.5%	60	*As % of total obs*		*79.6%*

Notes: Information on each country's official languages 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. Above tabulations by country of birth use following sample: 1990 IPUMS, arrived to the U.S. by age 17 between 1960 and 1974, is currently aged 25 to 38 and has non-missing value for English-speaking ability. "Countries" correspond to IPUMS detailed birthplace codes.

Appendix Table 2. Effect on Log Annual Wages -- Alternative Instruments

	first stage (1)	reduced-form (2)	OLS (3)	2SLS 2nd stage (4)	N (5)
Panel A. Base (from Table 3)					
English-speaking ability			0.2219 ***	0.3335 ***	47,422
(scale of 0 to 3, 3=best)			(0.0093)	(0.1054)	
max (0, age at arrival - 11) * non-English speaking country of birth	-0.0776 ***	-0.0259 ***			
	(0.0021)	(0.0082)			
Panel B. Linear Age at Arrival					
English-speaking ability			0.2219 ***	0.4519 ***	47,422
(scale of 0 to 3, 3=best)			(0.0093)	(0.1257)	
Age at arrival * non-English speaking country of birth	-0.0255 ***	-0.0115 ***			
	(0.0008)	(0.0032)			
Panel C. Dummy Variable for Arrival when Young					
English-speaking ability			0.2219 ***	0.4257 ***	47,422
(scale of 0 to 3, 3=best)			(0.0093)	(0.1218)	
(Age at arrival ≤ 11) * non-English speaking country of birth	0.2649 ***	0.1128 ***			
	(0.0084)	(0.0322)			
Panel D. All Three Instruments					
English-speaking ability			0.2219 ***	0.3571 ***	47,422
(scale of 0 to 3, 3=best)			(0.0093)	(0.1046)	
max (0, age at arrival - 11) * non-English speaking country of birth	-0.0627 ***	0.0003			
	(0.0039)	(0.0150)			
Age at arrival * non-English speaking country of birth	-0.0061 ***	-0.0071			
	(0.0011)	(0.0051)			
(Age at arrival ≤ 11) * non-English speaking country of birth	0.0156	0.0597			
	(0.0151)	(0.0599)			
Panel E. Age-at-Arrival Dummies					
English-speaking ability			0.2219 ***	0.3435 ***	47,422
(scale of 0 to 3, 3=best)			(0.0093)	(0.1045)	
Age-of-Arrival Dummies * non-English speaking country of birth	Yes	Yes			

Notes: Weighted by IPUMS weights. Robust standard errors in parentheses. Single asterisk denotes statistical significance at the 90% level of confidence, double 95%, triple 99%. Sample is as follows: 1990 IPUMS, arrived to the U.S. by age 17 between 1960 and 1974, is currently aged 25 to 38 and no missing data for wages and English-speaking ability. All specifications include age at arrival, country of birth, age, race/ethnicity and sex dummies.

UD035492

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC
Educational Resources Information Center

REPRODUCTION RELEASE
(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Language Skills and Earnings: Evidence from Childhood Immigrants

Author(s): Hoyt Bleakley and Aimee Chin

Corporate Source: Manuscript - Not yet published

Publication Date:

II. REPRODUCTION RELEASE:

Date of document is 10/1/2002

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to each document.

If permission is granted to reproduce and disseminate the identified documents, please CHECK ONE of the following three options and sign at the bottom of the page

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY SAMPLE TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY SAMPLE TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY SAMPLE TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
1	2A	2B
Level 1 ☑	Level 2A ☐	Level 2B ☐
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) and paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate these documents as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, please

Signature: [signature]

Printed Name/Position/Title: Aimee Chin / Assistant Prof.

Organization/Address: University of Houston, Dept. of Economics, 204 McElhinney Hall, Houston, TX 77204-5019

Telephone: 713 743 3761 FAX: 713 743 3798

E-Mail Address: achin@uh.edu Date: 2/18/2003

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of these documents from another source, please provide the following information regarding the availability of these documents. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:	Not Applicable
Address:	
Price:	

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:	Not Applicable
Address:	

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:	**ERIC Clearinghouse on Urban Education** **Teachers College Box 40, Columbia University** **525 West 120th Street** **New York, New York 10027-6696**

EFF-088 (Rev. 2/2000)