Negative and Positive Assimilation, Skill Transferability, and Linguistic Distance
Author(s): Barry R. Chiswick and Paul W. Miller
Source: *Journal of Human Capital*, Vol. 6, No. 1 (Spring 2012), pp. 35-55
Published by: The University of Chicago Press
Stable URL: http://www.jstor.org/stable/10.1086/664794

Negative and Positive Assimilation, Skill Transferability, and Linguistic Distance

Barry R. Chiswick
George Washington University and Institute for the Study of Labor (IZA)

Paul W. Miller
Curtin University

This paper synthesizes two models of immigrant assimilation: "positive assimilation" if earnings rise with duration as destination-relevant skills are acquired and "negative assimilation" if immigrants with highly transferable skills experience declining earnings as their economic rent diminishes. Hypotheses are developed and tested with earnings of adult male immigrants in the 2000 U.S. Census. "Linguistic distance" from English of an immigrant's mother tongue is the index of skill transferability. Only immigrants from English-speaking developed countries experience negative assimilation. Immigrants from other countries experience positive assimilation, the degree of assimilation increasing with linguistic distance.

I. Introduction

Research on immigrants over the past three decades has covered various issues, including their health, location, dominant language acquisition, marital and fertility patterns, use of public benefits, and labor market adjustment. The labor market adjustment literature has been based on a model that might be referred to as "positive assimilation" (Chiswick 1978, 1979). That is, immigrants move from a lower-income to a higher-income area, find that their premigration skills, including language skills, are not perfectly transferable, and engage in a process of investing in skills relevant for the destination, including destination-specific skills. These investments imply lower earnings in the investment period to be followed by increased earnings as they acquire skills more relevant for the destination. Conceptually and empirically, earnings increase, but at a decreasing rate, with duration in the destination. This model is con-

Miller acknowledges financial assistance from the Australian Research Council. Helpful comments have been received from two anonymous referees and participants at seminars at the Australian National University, Flinders University, and Macquarie University. We are grateful to Derby Voon and Marina Gindelsky for research assistance.

[*Journal of Human Capital*, 2012, vol. 6, no. 1]

sistent with data on immigrants' labor market adjustment for a wide range of destinations.

An alternative model of "negative assimilation" has recently been postulated (Chiswick and Miller 2011). In its simplest form, this model assumes that immigrants' origin and destination have similar levels of earnings and distribution of earnings, the skills are perfectly transferable between the origin and the destination, and, for simplicity, there are no investments in on-the-job training. This model implies that migration is a two-way street between the two countries. It will occur when a worker, searching in both countries in a global labor market, gets a much higher wage offer in the foreign country than in the origin. To the extent that this higher wage is above the norm for the worker's observable human capital (i.e., it is an economic rent), his earnings should decline with duration of residence in the destination. The negative assimilation model was tested by studying immigrants from the English-speaking developed countries (ESDCs) in the United States and Australia and Nordic immigrants to Sweden (Chiswick and Miller 2011). The hypothesized higher initial earnings than those of the native born, ceteris paribus, and the decline in earnings with duration of residence, were found to be consistent with the data. A study of changes in the hours of work of immigrants with duration of residence generates similar supportive evidence (see Blau, Kahn, and Papps 2011).

In the positive assimilation model, the rise in earnings with duration is attributable to skill and information acquisition. In the negative assimilation model, the decline in earnings is attributable to the decline in the economic rent that stimulated the initial migration.

If one had a measure of the transferability of skills between countries, what would be the skill transferability level that separated positive from negative assimilation? If both positive and negative assimilation can occur, what determines when positive assimilation becomes negative assimilation? To address these questions, one clearly needs a measure of skill transferability, preferably a scalar measure.

This paper analyzes the earnings assimilation of immigrants in the United States using an index of skill transferability to estimate the degree of skill transferability that determines whether assimilation is positive or negative. There is no obvious single measure, or even sets of measures, of the degree of skill transferability across countries. Among immigrants, proficiency in the destination language is an important skill, and countries with similar languages tend to have similar cultures and institutions. Indeed, the very high degree of linguistic transferability among the ESDCs and, to a slightly lesser extent, among the Nordic countries made these good test cases for the negative assimilation model.

Data from the 2000 U.S. Census on the earnings of adult male immigrants from the full range of origin countries are studied here to determine the level of skill transferability at which positive assimilation

(the rise in earnings with duration in the United States) becomes negative assimilation (the decline in earnings with duration in the United States). The key variable of interest is the interaction term between duration (years since migration) and a proxy measure of skill transferability (linguistic distance).

It is found that immigrants to the United States from ESDCs exhibit negative assimilation. Immigrants from all other countries, even those linguistically close to English, exhibit positive assimilation. Moreover, the degree of positive assimilation is positively related to the linguistic distance between English and the immigrants' language of origin. Tests of robustness support these findings.

Section II briefly summarizes the positive and negative assimilation models as well as the measure of "linguistic distance" from English of the mother tongues of immigrants and the alternative measures used in the tests of robustness. Section III describes the census data and provides the main empirical testing. It estimates the level of linguistic distance (skill transferability) at which the positive assimilation of immigrants turns to negative assimilation. Section IV presents, as tests of robustness, similar computations derived from estimations based on the percentage of the population in the country of origin that speaks English and on the English skills of recently arrived immigrants in the United States. Section V is a summary and conclusion.

II. Concepts

This section briefly summarizes the models of positive assimilation (Chiswick 1978, 1979), negative assimilation (Chiswick and Miller 2011), and the measure of the distance from English of the immigrant languages (Chiswick and Miller 2005). Details on the construction of the alternative measures of skill transferability used in the tests of robustness are also provided.

A. Positive Assimilation

Immigrants migrate from a low-wage country to a high-wage country. The migration goes in one direction. The immigrants arrive with skills, including language skills, which are not perfectly transferable. On arrival the immigrants have lower earnings than the native born, other measured variables the same, for two reasons. One is the less than perfect transferability to the destination of the skills acquired in the origin that they bring with them. The other is that earnings are reduced as the immigrants make investments in the destination to increase the transferability of previously acquired skills and to acquire new skills (including language skills). With the passage of time, earnings increase because of implementation of the newer modified skills and because the extent

of investments decreases.[1] That is, earnings increase at a decreasing rate with years since migration.

B. Negative Assimilation

On the other hand, consider two countries with equal levels of income and distribution of income and with perfect skill transferability between them.[2] Given that there are costs of migration, migration takes place only if the worker gets a job offer (for random or systematic reasons) that provides economic rent. Two-way migration occurs. In the destination the migrants initially have higher earnings than the native born, other things being the same, because otherwise they would not have moved. With the passage of time the economic rents dissipate; that is, earnings undergo a relative decline or a regression to the norm for the worker's observed set of human capital skills (Chiswick and Miller 2011). To the extent that migration costs are modest, an initial favorable wage offer, coupled with a fully anticipated regression toward the conditional mean, can still entice movements across national boundaries. The high initial earnings entice the move, but even with the regression toward the conditional mean, the earnings need not decline below what might have been earned in the origin.

Some of the immigrants will return to their origin. Others, however, stay in the destination because their earnings remain higher than in the origin or, if lower than in the origin, not sufficiently lower to encourage return migration. A factor inhibiting return migration is the acquisition of social ties and social capital in the destination. Marriage, children, and social networks in the destination all tend to discourage return migration. Among those who remain, a relative decline in earnings (regression to the conditional mean) with duration of residence would be expected. It is not that skills dissipate but that the economic rent that stimulated the initial move declines.

As argued in Chiswick and Miller (2011), there are other processes that would be consistent with a negative relationship between immigrants' earnings and their duration of residence in cross-sectional data. First, there could have been an increase over time in the unmeasured dimensions of the quality of immigrants. Second, the return migration that should be a feature of the data occurs predominately by the more successful in the destination country. This could arise where success in the destination helps attract even better wage offers in other countries. It is not obvious, however, why such positive selection should occur

[1] Labor market information networks are considered destination-specific skills in this context.

[2] While perfect skill transferability is assumed here, in practice all that is needed is a very high level of transferability. The empirical testing in this paper is designed to establish critical levels of such transferability.

among immigrants from the ESDCs that were analyzed in Chiswick and Miller (2011) but not among other groups of immigrants.

The estimation of a synthetic cohorts model by Chiswick and Miller (2011) indicated that cohort effects are of some, but not major, importance, and attempting to accommodate them empirically does not alter the conclusions in relation to either the presence of or reasons for negative assimilation. The review of the limited literature on return migration by Chiswick and Miller indicated that return migration was more likely to be characterized by negative selection than by positive selection. Negative selection would, ceteris paribus, lead to a positive relationship between earnings and duration of residence in analyses of cross-sectional data and, hence, serve to weaken any measured negative assimilation profile. Thus, the patterns documented below are unlikely to be due to these alternative hypotheses.

C. Skill Transferability

The primary measure of skill transferability used in this study is the "linguistic distance" from English of the immigrants' language of origin. This is based on the assumption that language skills play a direct role in the labor market in the job investment process, in the job search process, and on the job. It is further assumed that a language more distant from English represents a less transferable skill. Chiswick and Miller (2005) developed a measure of linguistic distance based on the difficulty that Americans have learning the language.[3] It is assumed that if Americans have more difficultly learning a specific language, native speakers of the language have more difficulty learning English.

On the basis of the tests given to Americans studying a limited range of foreign languages, linguistic scores were established that range from 1.0 for languages most distant from English (Japanese and Korean) to 3.0 for languages closest to English (Afrikaans, Norwegian, and Swedish) (Hart-Gonzalez and Lindermann 1993). On the basis of the closeness of languages, linguistic scores were established for nearly all of the remaining languages (other than native American Indian languages and a few languages of unknown origin; see Grimes and Grimes 1993; Chiswick and Miller 2005). The linguistic scores for the 43 most frequent

[3] The linguistic distance measure was created for nearly all of the languages coded in the U.S. 1990 and 2000 censuses except for the languages of the native peoples of the Americas and a few unique languages (e.g., Basque). Few of these speakers would be among the immigrants.

languages spoken by immigrants in the United States are reported in Appendix table A1.[4]

Three alternatives to the measure of linguistic distance are considered. The first is the percentage of people in the immigrant's country of origin who speak English. There are various sources for these data, but disparate sources raise concerns about comparability. Accordingly, only data compiled in two sources are used. First, Crystal (2003) provides relevant data for over 75 countries or territories, many of which were former British colonies. These data cover both individuals who speak English as a first language and those for whom English is a second language. Second, the European Commission (2006, 13) collected data in 2005 for 25 EU countries as well as for Bulgaria, Romania, Croatia, and Turkey. This alternative assumes that migrants are drawn from the linguistic majority in the origin country.

The second alternative is a dichotomous variable, set equal to one if English is one of the country's official languages. These data were extracted from Banks (1988). The third proxy for the degree of skill transferability is based on country-level data on the English skills of recently arrived immigrants in the United States. These data refer to the incidence of English speaking proficiency among adult immigrants with 5 or fewer years of residence in the United States at the time of the 2000 Census.[5] A limitation of this measure is that it has a focus only on the most recent arrival cohort, and the selection of immigrants may have changed over time.

For each of these alternatives the assumption is that countries with English as an official language and countries where a greater proportion of the population speak English will provide a setting in which human capital can be accumulated that is more readily transferable to the labor market of the United States. The absence of information on English skills at the time of arrival in the census data used in Sections III and IV precludes consideration of further measures.

III. The Data and Empirical Analysis

The empirical testing of the model presented below is based on the 2000 U.S. Census of Population, 1 percent Public Use Microdata Sample for adult (aged 25–64) foreign-born males with earnings. Earnings are the sum of wage, salary, and self-employment income in 1999. As is

[4] These linguistic scores have been used to study immigrant earnings as well as international trade patterns (Chiswick and Miller 2005; Hutchinson 2005). Van der Slik (2010) provides discussion of these scores, along with alternatives. Van der Slik is critical of the symmetry assumption (i.e., if it is easy for a native English speaker to learn foreign language *X*, then it is equally easy for a native speaker of foreign language *X* to learn English) of Chiswick and Miller (2005), though this is necessary to develop this practical proxy of the linguistic distance between English and a wide range of other languages.

[5] The 5 percent Public Use Microdata Sample was used to extract these data. Both male and female immigrants are considered.

standard in research on immigrant earnings, the natural logarithm of annual earnings is regressed on years of schooling; years since migration to the United States;[6] years of potential labor market experience and its square; dichotomous variables for marital status (married spouse present, married = 1), race (black = 1), and location (South = 1, metropolitan area = 1); and the natural logarithm of weeks worked (ln *WW*). Tests show that the main findings do not change if additional controls (e.g., extra race variables) are included in the estimating equation. Hence only the findings from this parsimonious model are presented.

The language question in the 2000 Census asks if the person speaks a language other than English at home. If the response is in the affirmative, the person is then asked to identify this language and to self-report how well this person speaks English: very well, well, not well, or not at all. With those who speak only English at home, this creates five English proficiency categories. In the regression analysis, those who speak a language other than English at home and speak English very well or who speak it well are combined into one category, as are those who speak English not well or not at all. The benchmark group is those who speak only English at home.

The self-reported non-English languages spoken at home are identified in the 2000 Census microdata file and codebook. Except for Native American languages, there is a measure of linguistic distance from English for nearly all the other languages identified in the census microdata language list (Chiswick and Miller 2005). The measure used here is referred to as the linguistic score. The alternative measures of skill transferability used in the tests of robustness in Section IV are based on the English-language characteristics of the worker's country of birth.

Each immigrant from a non-English-speaking developed country (non-ESDC) who reports a non-English language is assigned the relevant value for the linguistic score. If the immigrant from a non-ESDC reports that he speaks only English, the mean language score for immigrants from that country who speak a language other than English is assigned.

In the first estimating equation, immigrants to the United States from the ESDC (Australia, Canada, Ireland, New Zealand, and the United Kingdom) are treated as one group and have a separate variable for years since migration, YSM (ESDC). All other countries are first constrained to have the same partial effect of duration in the United States (years since migration) on earnings, YSM (non-ESDC). This separate years since migration variable is then interacted with the linguistic score. For immigrants from the ESDCs, which is abbreviated to *E* in the regression coefficient that follows, the postmigration earnings adjustment

[6] A linear years since migration variable is used in preference to a quadratic as the linear model generates the same substantive findings as the more general model but provides a basis for a clearer presentation of the research findings when multiple interaction terms are involved.

is given by $\partial \ln Y/\partial \text{YSM(ESDC)} = \alpha^{E}$, which is hypothesized to be negative under the negative assimilation model. For immigrants from non-ESDC countries (abbreviated to *NE* in the regression coefficients that follow), the postmigration growth in earnings is given by

$$\frac{\partial \ln Y}{\partial \text{YSM(non-ESDC)}} = \alpha_1^{NE} + \alpha_2^{NE}(\text{LS}),$$

where α_1^{NE} is hypothesized to be positive and α_2^{NE} is hypothesized to be negative as the linguistic score (LS) is higher the closer the language is to English.

The main hypothesis investigated is whether immigrants from countries with languages close to English (high linguistic score) have a flatter or even negative profile for earnings with respect to the duration of residence. This framework can be used to determine the value of the linguistic score at which the postimmigration change in earnings for those from non-ESDCs would be zero, representing neither positive nor negative assimilation. In other words, at what value, if any, of LS does $\alpha_1^{NE} + \alpha_2^{NE}(\text{LS}) = 0$. This hypothetical is asking where the linguistic score would fall if a linear scale were to be used to assess the language for which there is neither positive nor negative assimilation.

Table 1 reports the relevant regression coefficients, with the full equation reported in Appendix B (see table B1).[7] The samples in columns 1 and 2 pertain to all immigrants. The next three columns pertain to immigrants from the non-ESDCs partitioned according to the distance of their mother tongue from English (from most distant, col. 3, to intermediate values, col. 4, to closest to English, col. 5), and column 6 pertains to immigrants from the ESDCs.

In table 1, column 1, there is a dichotomous variable for ESDCs and a variable for the effects of duration in the United States for immigrants from these countries, YSM (ESDC). Ceteris paribus, earnings are substantially higher for these immigrants (coefficient 0.53, *t*-ratio = 12.4). Consistent with Chiswick and Miller (2011), earnings decrease among these immigrants by about one-half of a percentage point per year since migration (coefficient −0.005, *t*-ratio = 4.75). Among immigrants from other countries, however, earnings increase with duration in the United States, YSM (non-ESDC). The positive effect of duration becomes smaller the larger the linguistic score, that is, the closer the origin language is to English. But how close does the linguistic score have to be for there to be neither positive nor negative assimilation? The regression analysis in table 1 (full equation in table B1) implies a score

[7] As is standard in native-born and immigrant earnings equations, in table B1, annual earnings increase with educational attainment, total labor market experience, weeks worked, not being racially black, being currently married, living in a metropolitan area, living outside the South, and being more proficient in English. As these are standard findings, they are not discussed further.

TABLE 1
Selected Regression Results from Analysis of Immigrant Earnings with Linguistic Distance Variable, Adult Males, 2000 U.S. Census

	Model					
Variable	(1)	(2)	(3)	(4)	(5)	(6)
ESDC	.534	*	*	*	*	*
	(12.37)					
YSM (ESDC)	−.0046	−.0044	*	*	*	−.0052
	(4.75)	(4.54)				(4.76)
YSM (non-ESDC)	.0148	*	.0130	.0085	.0052	*
	(8.21)		(19.06)	(20.86)	(5.31)	
Linguistic score	−.0095	*	*	*	*	*
	(.57)					
YSM (non-ESDC) × linguistic score	−.0027	*	*	*	*	*
	(3.22)					
LS1 (far from English)	*	−.578	*	*	*	*
		(21.65)				
LS2 (intermediate)	*	−.610	*	*	*	*
		(23.85)				
LS3 (close to English)	*	−.401	*	*	*	*
		(13.17)				
LS1 × YSM (non-ESDC)	*	.012	*	*	*	*
		(20.71)				
LS2 × YSM (non-ESDC)	*	.009	*	*	*	*
		(22.14)				
LS3 × YSM (non-ESDC)	*	.007	*	*	*	*
		(8.86)				
Adjusted R^2	.363	.366	.325	.349	.351	.266
Sample size	84,052	84,052	23,896	48,009	7,802	4,345

Source.—Extracted from table B1 on the basis of the 2000 U.S. Census 1 percent Public Use Microdata Sample.
Note.—Heteroskedasticity-consistent *t*-statistics are in parentheses. Columns 1 and 2 pertain to the entire sample; cols. 3–5 pertain to the linguistic score groups LS1, LS2, and LS3 from the non-ESDCs; and col. 6 pertains to immigrants from the ESDCs.
* The variable is not included.

of 5.5.[8] Yet this is outside the range of the data, as the highest linguistic score is 3.0 for the languages closest to English (primarily for Swedish- and Norwegian-speaking immigrants).

The linguistic score variable in the column 1 specification is not statistically significant. This is a consequence of the inclusion of the interaction term with years since migration. In models that exclude such an interaction term, the linguistic score is a statistically significant determinant of earnings, and the estimated coefficient indicates that the earnings of immigrants with languages close to English exceed those of

[8] Evaluating $\partial \ln Y/\partial \text{YSM(non-ESDC)}$, the term $0.0148 - 0.0027(\text{LS})$ becomes zero at approximately LS = 5.5. The linguistic score range is from 1.0 to 3.0. Note that we have only a mean score for each mother tongue. If we had the underlying individual test scores for each mother tongue, we could assess where on the right-hand tail of the distribution of test scores the score of 5.5 would be. Presumably, however, it would represent few workers.

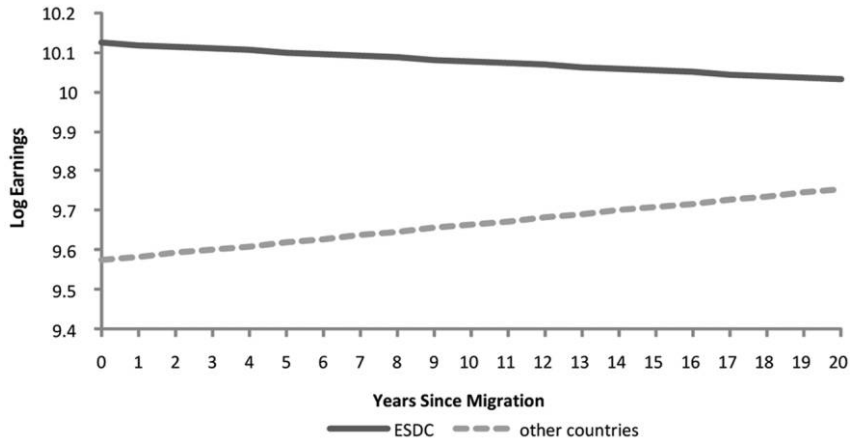

Figure 1.—Patterns of postarrival earnings adjustment for immigrants in the United States by region of origin. Source: Authors' calculations from column 1 of table 1.

immigrants with languages more distant from English (see Chiswick and Miller 1998).

In table 1, column 2, the linguistic score term and its interaction with duration are replaced by dichotomous linguistic score variables for languages far from English (LS1), at an intermediate distance from English (LS2), and close to English (LS3). These variables are relevant only for immigrants from the non-ESDCs and form a mutually exclusive and exhaustive set of dichotomous variables for this part of the sample. Earnings are lower for the LS1 and LS2 languages than for the languages closest to English (LS3).[9] While the negative effect of the linguistic score for these languages diminishes with duration in the United States, the effect never disappears.

In table 1, columns 3–5, separate equations are computed for the three non-English linguistic groups from non-ESDCs. The partial effect of duration on earnings is most positive for the languages (LS1) most distant from English (coefficient 0.013, *t*-ratio = 19.1), followed by the intermediate languages (coefficient 0.009, *t*-ratio = 20.9), with the languages closest to English having a smaller but still highly significant positive effect of duration on earnings (coefficient 0.005, *t*-ratio = 5.3). It is only in table 1, column 6, where the analysis is limited to immigrants from the ESDCs, that the effect of duration is negative and statistically significant (coefficient −0.005, *t*-ratio = 4.8).

Figure 1 depicts the contrasting postarrival pattern of earnings adjustment for immigrants from the ESDCs and for immigrants from other

[9] The relatively low earnings of LS2 speakers may be due to the inclusion of Spanish, as Hispanics tend to have lower earnings than other immigrants, other variables being the same.

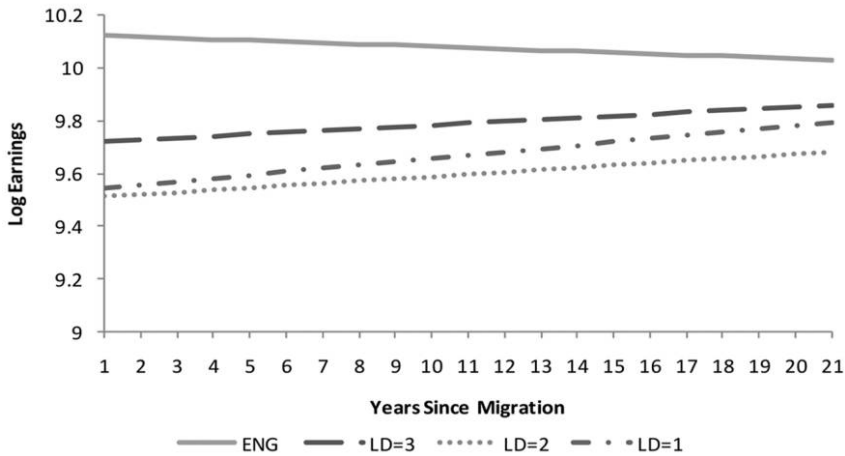

Figure 2.—Patterns of postarrival earnings adjustment for immigrants in the United States by region of origin and linguistic distance from English. Source: Authors' calculations from column 2 of table 1.

countries. The figure shows the much higher initial earnings of the ESDC immigrants, other variables the same, and the narrowing of the gap as time in the United States increases. Earnings decrease with duration for ESDC immigrants and increase with duration for other immigrants.

Figure 2 shows the pattern of earnings with respect to duration for immigrants from the ESDCs and the three separate linguistic groups among other immigrants. Initial earnings are lowest for the language groups furthest from English, but the increase with duration is steeper the further the language is from English. This demonstrates the by now familiar pattern of postarrival earnings growth varying inversely with the initial postarrival earnings (Chiswick 1979). This analysis shows the crucial role played by the linguistic distance of the immigrants' mother tongue from English in generating this pattern.

IV. Tests of Robustness

The findings above appear to be convincing. However, they are based on a single index of skill transferability provided by the linguistic scores of Hart-Gonzalez and Lindermann (1993). To examine the robustness of the findings, the analyses are first replicated using the percentage of the population in the immigrants' country of origin who speak English as the index of skill transferability.[10] To accommodate the fact that this

[10] Blau et al. (2011) use source country characteristics in their study of the labor market assimilation among immigrants in the United States.

index is available on a consistent basis for only a subset of countries that are not ESDCs, the following variables are added to the estimating equation:

- the percentage who speak English in the non-ESDC country of origin;
- a dummy variable for non-ESDCs without the requisite data on the above variable;
- separate years since migration terms for immigrants from the non-ESDCs with and without the information on the percentage who speak English;
- interaction of the percentage who speak English in the country of origin and years since migration.

It will be apparent that the first and final variables listed above take the place of the linguistic score and linguistic score × years since migration interactions of table 1. The variables mentioned in the second and third points provide a generalization of the linear years since migration variable for immigrants from non-ESDCs used previously, which is necessitated by the additional birthplace categorization utilized in the current set of analyses.

Table 2 reports the relevant regression estimates from this alternative set of analyses. Column 1 pertains to the full sample and is similar to column 1 in table 1. It has five main features. First, the earnings effects for immigrants from ESDCs are broadly the same as those estimated in table 1. Thus, these immigrants are associated with a substantial ceteris paribus earnings advantage at arrival, and their earnings decrease with length of stay in the United States. Second, the variable for the percentage who speak English is statistically insignificant. It has a coefficient of −0.0005 and a *t*-value of 1.42. Similarly, the linguistic score in table 1 was not a statistically significant influence on earnings. Third, immigrants from countries for which there are data on the percentage who speak English have higher earnings than immigrants from the other non-ESDCs. Fourth, the baseline (for a zero percentage who speak English) effect of years since migration on earnings for immigrants from the non-ESDCs with data on the percentage who speak English (as noted previously, these countries are primarily former British colonies and EU countries) is similar to that for the non-ESDCs, for which such data are not available. The relevant estimates for these two groups of countries are 0.0106 ($t = 27.56$) and 0.0100 ($t = 9.92$), respectively. Fifth, the years since migration effect is negatively related to the percentage who speak English in the immigrant's country of origin. The interaction term between the percentage who speak English and the years since migration variable is −0.000067 ($t = 3.50$). Thus, this pattern of effects is identical to that reported above on the basis of the linguistic score. Similarly to the situation in which the linguistic score is used, there is no relevant

TABLE 2
SELECTED REGRESSION RESULTS FROM ANALYSIS OF IMMIGRANT EARNINGS WITH ALTERNATIVE MEASURES OF SKILL TRANSFERABILITY, ADULT MALES, 2000 U.S. CENSUS

	Model					
Variable	(1)	(2)	(3)	(4)	(5)	(6)
ESDC	.445	*	*	*	*	.603
	(13.70)					(23.70)
YSM (ESDC)	−.0044	*	*	*	−.0052	−.0045
	(4.54)				(4.76)	(4.67)
non-ESDC (no data on %ES)	−.248	*	*	*	*	*
	(12.35)					
YSM (non-ESDC, no data on %ES)	.0106	.0105	*	*	*	*
	(27.56)	(26.25)				
%ES	−.0005	*	*	.0007	*	*
	(1.42)			(1.67)		
YSM (non-ESDC, with data on %ES)	.0100	*	.0079	.0124	*	*
	(.92)		(13.30)	(11.39)		
YSM (non-ESDC, with data on %ES) × %ES/100	−.0067	*	*	−.0098	*	*
	(3.50)			(5.01)		
English as OFFL	*	*	*	*	*	.1573
						(10.27)
YSM × (non-ESDC with OFFL)	*	*	*	*	*	.0077
						(9.72)
YSM (non-ESDC without OFFL)	*	*	*	*	*	.0099
						(28.39)
Adjusted R^2	.368	.345	.318	.319	.266	.364
Sample size	84,052	59,660	20,047	20,047	4,345	84,052

Source.—Extracted from table B2, on the basis of the 2000 U.S. Census 1 percent Public Use Microdata Sample.
Note.—Heteroskedasticity-consistent *t*-statistics are in parentheses. Columns 1 and 6 pertain to the entire sample; col. 2 pertains to immigrants from non-ESDCs where there are no data on the percentage speaking English (%ES); cols. 3 and 4 pertain to immigrants from countries having data on %ES; and col. 5 pertains to immigrants from the ESDCs. OFFL = official language.
* The variable is not included.

value of the variable for the percentage who speak English where positive assimilation turns into negative assimilation.[11]

Columns 2–5 of table 2 present results for various subsets of the data. Column 2 lists results from the model estimated on immigrants from non-ESDCs for which there are no data on the percentage who speak English. The years since migration effect for this group of countries is 0.0105 ($t = 26.3$), which is statistically indistinguishable from the impact estimated (0.0106) in the pooled sample in column 1.

Columns 3 and 4 of table 2 list results for the group of countries for which there are data on the percentage of the population who speak English. The pattern that emerges from these data is that the years since migration effect is 0.0124 for the baseline case in which a zero per-

[11] The estimates imply that this threshold occurs at a proportion of 1.49, when, by definition, the variable ranges in value from 0.0 to 1.0

centage speaks English. This effect dissipates at the rate of 0.0001 for each percentage point increase in the proportion of English speakers in the country of origin. Hence, for a country where all the population speaks English, the years since migration effect will be a rather minuscule 0.26 percent increase in earnings with each extra year of stay in the United States ($0.26 = [0.0124 - 0.000098] \times 100$). Thus, for a country that is not one of our ESDCs, even if the entire population is reported to speak English, the immigrants to the United States experience a small positive assimilation in their earnings. Column 5 presents, for comparison purposes, the results for immigrants from ESDCs. These results are the same as in column 5 of table 1.

Column 6 of table 2 lists results from an alternative specification, where the non-ESDCs are categorized into two: countries with English as an official language and other countries. Countries with English as an official language are associated with 15.73 percent higher earnings. The years since migration effect for this group of countries is a statistically significant 0.0077 ($t = 9.7$). While positive, this estimate is significantly lower than the estimated years since migration effect of 0.0099 ($t = 28.4$) for the other non-ESDCs. These results suggest that greater insights are derived when continuous proxies for the degree of skill transferability are employed, such as those provided by the linguistic score and the percentage of English speakers of the population in the country of origin.

This appraisal is reinforced through the use of an alternative index of skill transferability created from the 2000 Census data. The incidence of speaking English only or where a language other than English was spoken at home, the incidence of speaking English very well or well (denoted "good English" below), was computed for each country of origin among adult immigrants who had resided in the United States for 5 or fewer years. This index was included in the estimating equation in place of the percentage who speak English in the country of origin. The results are entirely consistent with those obtained when the percentage who speak English in the country of origin variable was used. In particular, the estimates on the new variables (for good English and its interaction with the duration of residence in the United States) were

$$\underset{(0.69)}{0.100}\text{good English} + \underset{(16.70)}{0.011}\text{YSM} - \underset{(3.27)}{0.002}\text{YSM} \times \text{good English}.$$

Thus, these estimates show that there are higher initial earnings and a weaker rise in earnings with duration among immigrants from countries that appear to be characterized by greater skill transferability. The negative effect of the impact of duration (YSM) on earnings associated with the good English variable is quite diluted compared to that derived when characteristics of the countries of origin are used. This suggests that the data on immigrants in a country (e.g., good English) are a poor substitute for the origin country characteristics (percentage speak-

ing English in the country of origin). Alternatively stated, selection in migration appears to dilute the true influence of the country of origin characteristics.

V. Summary and Conclusion

The conceptual framework used in the study of immigrants' social and economic adjustment was broadened in Chiswick and Miller's (2011) recent paper, where a process of negative assimilation in the postarrival earnings was proposed to sit alongside the conventional positive assimilation model. Negative assimilation was developed and tested in the context of immigrants from countries that are very similar to the host country in terms of the transferability of skills, culture, and labor market institutions. Specifically, the analyses were primarily based on immigrants migrating from English-speaking developed countries to another English-speaking developed country in response to a favorable draw from the earnings distribution in the destination country.

In the positive assimilation model, earnings increase with duration of residence because of the accumulation of skills, including knowledge, relevant for the destination labor market. In the negative assimilation model, earnings decrease with duration because the economic rent that stimulated the migration decreases over time.

The current study has generalized the conceptual framework behind the negative assimilation hypothesis to immigrants in the United States through analyzing the postarrival earnings profiles of immigrants according to the linguistic distance of their mother tongue from English. The findings show that immigrants from non-ESDCs are characterized by positive assimilation. The extent of this positive assimilation varies, however, with the linguistic distance of their mother tongue from English. The positive earnings effect associated with duration of residence in the United States is less intense for immigrants with a mother tongue closer to English than it is for immigrants with a mother tongue more distant from English. This pattern of effects was established using various alternative specifications of the estimating equation, namely, a model with a linear linguistic distance variable interacted with the years since migration variable and a model based on three dichotomous variables formed for separate groups of immigrants on the basis of their value on the linguistic distance measure. Similar patterns were established with alternative measures of the transferability of origin country human capital. In other words, the findings are robust to the measure used.

Immigrants' postarrival earnings growth varies according to the similarity of their background characteristics to those of the native born in the host country. Among immigrants, earnings in the immediate postarrival period are higher for those with a mother tongue closer to English and higher still for immigrants from the ESDCs. Postarrival growth in earnings is, however, greater for the group with a mother tongue

more distant from English. It is positive but more modest for groups with a non-English mother tongue that is closer to English and negative for the limiting case of immigrants from the ESDCs.

The analysis of the earnings of immigrants can be used to compute the value of the linguistic score that would result in neither negative nor positive assimilation, that is, no effect of duration on earnings. This value is beyond the range of the linguistic scores for the non-English languages. Even the language groups closest to English—but not English—exhibit positive assimilation.

Estimates of models of immigrant assimilation that eschew information on immigrants' heterogeneity with respect to their mother tongue will therefore hide important aspects of the initial level and postarrival growth in earnings. Whether these patterns hold for other indices of country of origin differences (e.g., in institutions or workplace cultures) is a topic for further research.

Appendix A

Definitions of Variables

The variables used in the statistical analyses are defined below.

Data source: 2000 Census of Population, Public Use Microdata Sample, 1 percent sample.

Definition of population: 25–64-year-old foreign-born males with positive earnings in 1999 from wages and salaries or self-employment.

Dependent Variable

Earnings in 1999: Natural logarithm of the annual earnings in 1999 from wages and salaries and self-employment income.

Explanatory Variables

Educational attainment: This variable records the total years of full-time-equivalent education. It has been constructed from the census data on educational attainment by assigning the following values to the census categories: completed less than fifth grade (2 years); completed fifth or sixth grade (5.5); completed seventh or eighth grade (7.5); completed ninth grade (9); completed tenth grade (10); completed eleventh grade (11); completed twelfth grade or high school (12); attended college for less than 1 year (12.5); attended college for more than 1 year or completed college (14); bachelor's degree (16); master's degree (17.5); professional degree (18.5); doctorate (20).

Experience: age − years of education − 6.

Weeks worked: This variable is the natural logarithm of the weeks worked in 1999.

Black: This is a dichotomous variable that distinguishes immigrants of self-reported black racial origin from all other racial origins.

TABLE A1
Linguistic Score: Index of Difficulty of Learning English by Mother Tongue

Language	Language Score	Language	Language Score
L3:		L1 (*continued*)	
Afrikaans	3.00	Hebrew	2.00
Norwegian	3.00	Hungarian	2.00
Rumanian	3.00	Indonesian	2.00
Swedish	3.00	Mongolian	2.00
Dutch	2.75	Polish	2.00
Malay	2.75	Serbo-Croatian	2.00
Swahili	2.75	Tagalog	2.00
French	2.50	Thai	2.00
Italian	2.50	Turkish	2.00
Portuguese	2.50	Bengali	1.75
L2:		Burmese	1.75
Danish	2.25	Greek	1.75
German	2.25	Hindi	1.75
Russian	2.25	Nepali	1.75
Spanish	2.25	Sinhala	1.75
L1:		Arabic	1.50
Amharic	2.00	Lao	1.50
Bulgarian	2.00	Mandarin	1.50
Cambodian	2.00	Vietnamese	1.50
Czech	2.00	Cantonese	1.25
Dari	2.00	Japanese	1.00
Farsi	2.00	Korean	1.00
Finnish	2.00		

Source.—Chiswick and Miller (2005), table 1.
Note.—A higher score means less difficulty in learning English. LS1: far from English, linguistic score 2.0 or below; LS2: intermediate distance, linguistic score greater than 2.0 and less than or equal to 2.25; LS3: close to English, linguistic score greater than 2.25.

Marital status: This is a dichotomous variable that distinguishes individuals who are married, spouse present (equal to one), from all other marital states.

Location: The two dichotomous location variables record residence in a metropolitan area or a southern state.

English proficiency: There are two dichotomous variables for self-reported English proficiency. The first distinguishes individuals who speak a language other than English at home and speak English either very well or well. The second is for individuals who speak a language other than English at home and either speak English not well or speak English not at all. The benchmark group is those who speak only English at home.

English-speaking developed countries: The ESDCs considered in this study are the United Kingdom, Ireland, Canada, Australia, and New Zealand.

Years since migration: This is computed from the year the foreign-born person came to the United States to stay.

Linguistic distance: See table A1, explained in Chiswick and Miller (2005). The dummy variables LS1, LS2, and LS3 refer to language scores for immigrants from non-ESDCs and are as follows: LS1: far from English, LS 2.0 or below; LS2: intermediate distance, LS greater than 2.0 and less than or equal to 2.25; LS3: close to English, LS values over 2.25.

Percentage who speak English in the country of origin: a continuous variable, derived from Crystal (2003) and the European Commission (2006). This is relevant/available only for a subset of the non-ESDCs.

English as an official language: a dichotomous variable, set equal to one where English is an official language in the non-ESDC country of origin and set equal to zero otherwise. These data are from Banks (1988).

Good English: This is a continuous variable, defined as the percentage of immigrants who have resided in the United States for 5 or fewer years who are proficient in English.

Appendix B

TABLE B1
Regression Results from Analysis of Immigrant Earnings with Linguistic Distance Variable, Adult Males, 2000 U.S. Census

	Model					
Variable	(1)	(2)	(3)	(4)	(5)	(6)
Constant	5.775	6.369	5.340	5.986	5.421	4.374
	(87.69)	(106.55)	(43.48)	(92.71)	(28.54)	(13.81)
Educational attainment	.053	.050	.074	.040	.055	.111
	(67.54)	(62.38)	(37.51)	(42.74)	(17.69)	(19.62)
Experience (EXPER)	.010	.010	.010	.008	.011	.049
	(9.31)	(9.14)	(4.79)	(5.98)	(3.35)	(9.19)
$EXPER^2/100$	−.015	−.016	−.023	−.012	−.018	−.082
	(7.78)	(8.30)	(5.42)	(4.75)	(2.95)	(7.61)
Log weeks worked	.875	.874	.920	.830	1.004	.992
	(73.19)	(73.02)	(37.13)	(56.68)	(24.21)	(15.01)
Married	.211	.211	.190	.211	.235	.251
	(35.56)	(35.47)	(14.64)	(29.54)	(11.89)	(8.22)
Black	−.115	−.179	−.272	.007	−.299	−.169
	(10.54)	(14.92)	(5.87)	(.42)	(14.42)	(2.54)
South	−.069	−.062	−.059	−.054	−.123	−.069
	(11.37)	(10.17)	(4.39)	(7.42)	(6.04)	(2.24)
Metropolitan	.137	.130	.148	.124	.183	.279
	(5.10)	(4.85)	(2.03)	(4.58)	(2.05)	(1.57)
Speaks English very well/well	−.060	−.039	−.086	.003	−.045	−.024
	(6.25)	(4.06)	(4.11)	(.18)	(2.25)	(.64)
Speaks English not well/not at all	−.251	−.216	−.256	−.186	−.122	−.133
	(22.44)	(19.07)	(9.44)	(12.55)	(2.91)	(.87)
ESDC	.534	*	*	*	*	*
	(12.37)					
YSM (ESDC)	−.0046	−.0044	*	*	*	−.0052
	(4.75)	(4.54)				(4.76)
YSM (non-ESDC)	.0148	*	.0130	.0085	.0052	*
	(8.21)		(19.06)	(20.86)	(5.31)	
Linguistic score	−.0095	*	*	*	*	*
	(.57)					
YSM (non-ESDC) × linguistic score	−.0027	*	*	*	*	*
	(3.22)					
LS1 (far from English)	*	−.578	*	*	*	*
		(21.65)				
LS2 (intermediate)	*	−.610	*	*	*	*
		(23.85)				

TABLE B1 (*Continued*)

Variable	Model (1)	(2)	(3)	(4)	(5)	(6)
LS3 (close to English)	*	−.401	*	*	*	*
		(13.17)				
LS1 × YSM (non-ESDC)	*	.012	*	*	*	*
		(20.71)				
LS2 × YSM (non-ESDC)	*	.009	*	*	*	*
		(22.14)				
LS3 × YSM (non-ESDC)	*	.007	*	*	*	*
		(8.86)				
Adjusted R^2	.363	.366	.325	.349	.351	.266
Sample size	84,052	84,052	23,896	48,009	7,802	4,345

Source.—2000 U.S. Census 1 percent Public Use Microdata Sample.
Note.—Heteroskedasticity-consistent *t*-statistics are in parentheses. Columns 1 and 2 pertain to the entire sample; cols. 3–5 pertain to linguistic score groups LS1, LS2, and LS3 for immigrants from non-ESDCs; and col. 4 pertains to immigrants from ESDCs.
* The variable is not included.

TABLE B2
Regression Results from Analysis of Immigrant Earnings with Alternative Measures of Skill Transferability, Adult Males, 2000 U.S. Census

Variable	Model (1)	(2)	(3)	(4)	(5)	(6)
Constant	5.973	5.865	5.355	5.347	4.374	5.717
	(103.57)	(97.27)	(42.10)	(41.55)	(13.81)	(105.52)
Educational attainment	.050	.045	.074	.074	.111	.052
	(63.57)	(52.42)	(34.10)	(33.72)	(19.62)	(67.56)
Experience (EXPER)	.011	.013	−.002	−.003	.049	.010
	(10.08)	(9.81)	(1.17)	(1.38)	(9.19)	(9.73)
$EXPER^2/100$	−.018	−.020	.003	.004	−.082	−.017
	(9.42)	(8.94)	(.75)	(.91)	(7.61)	(8.48)
Log weeks worked	.874	.842	.973	.971	.992	.876
	(73.08)	(62.60)	(35.05)	(35.01)	(15.01)	(73.20)
Married	.208	.187	.266	.266	.251	.210
	(35.12)	(27.78)	(20.09)	(20.07)	(8.22)	(35.28)
Black	−.132	−.097	−.177	−.145	−.169	−.175
	(11.20)	(6.34)	(10.51)	(7.54)	(2.54)	(15.56)
South	−.061	−.057	−.093	−.093	−.069	−.067
	(9.98)	(8.37)	(6.67)	(6.69)	(2.24)	(11.05)
Metropolitan	.129	.125	.123	.122	.279	.134
	(4.83)	(4.59)	(1.98)	(1.97)	(1.57)	(5.02)
Speaks English very well/well	−.044	−.008	−.043	−.064	−.024	−.040
	(4.23)	(.57)	(2.93)	(3.95)	(.64)	(4.11)
Speaks English not well/not at all	−.205	−.196	−.130	−.147	−.133	−.216
	(16.98)	(12.64)	(4.29)	(4.67)	(.87)	(18.84)
ESDC	.445	*	*	*	*	.603
	(13.70)					(23.70)
YSM (ESDC)	−.0044	*	*	*	−.0052	−.0045
	(4.54)				(4.76)	(4.67)
Non-ESDC (no data on %ES)	−.248	*	*	*	*	*
	(12.35)					
YSM (non-ESDC, no data on %ES)	.0106	.0105	*	*	*	*
	(27.56)	(26.25)				

TABLE B2 (*Continued*)

Variable	Model (1)	(2)	(3)	(4)	(5)	(6)
%ES	−.0005 (1.42)	*	*	.0007 (1.67)	*	*
YSM (non-ESDC, with data on %ES)	.0100 (9.92)	*	.0079 (13.30)	.0124 (11.39)	*	*
YSM (non-ESDC, with data on %ES) × %ES/100	−.0067 (3.50)	*	*	−.0098 (5.01)	*	*
English as OFFL	*	*	*	*	*	.1573 (10.27)
YSM × (non-ESDC with OFFL)	*	*	*	*	*	.0077 (9.72)
YSM × (non-ESDC without OFFL)	*	*	*	*	*	.0099 (28.39)
Adjusted R^2	.368	.345	.318	.319	.266	.364
Sample size	84,052	59,660	20,047	20,047	4,345	84,052

Source.—2000 U.S. Census 1 percent Public Use Microdata Sample.
Note.—Heteroskedasticity-consistent *t*-statistics are in parentheses. Columns 1 and 6 pertain to the entire sample; col. 2 pertains to immigrants from non-ESDCs where there are no data on the percentage speaking English (%ES); cols. 3 and 4 pertain to immigrants from countries having data on %ES; and col. 5 pertains to immigrants from ESDCs.
* The variable is not included.

References

Banks, Arthur S. 1988. *Political Handbook of the World 1988*. Binghamton, NY: CSA Pubs.

Blau, Francine D., Lawrence M. Kahn, and Kerry L. Papps. 2011. "Gender, Source Country Characteristics and Labor Market Assimilation among Immigrants: 1980–2000." *Rev. Econ. and Statis.* 93 (1): 43–58.

Chiswick, Barry R. 1978. "The Effect of Americanization on the Earnings of Foreign-Born Men." *J.P.E.* 86 (5): 897–921.

———. 1979. "The Economic Progress of Immigrants: Some Apparently Universal Patterns." In *Contemporary Economic Problems, 1979*, edited by William Fellner, 357–99. Washington, DC: American Enterprise Inst.

Chiswick, Barry R., and Paul W. Miller. 1998. "English Language Fluency among Immigrants in the United States." In *Research in Labor Economics*, vol. 17, edited by Solomon W. Polachek, 151–200. Stamford, CT: JAI.

———. 2005. "Linguistic Distance: A Quantitative Measure of the Distance between English and Other Languages." *J. Multilingual and Multicultural Development* 26 (1): 1–11.

———. 2011. "The 'Negative' Assimilation of Immigrants: A Special Case." *Indus. and Labor Relations Rev.* 64 (3): 502–25.

Crystal, D. 2003. *The Cambridge Encyclopedia of the English Language*. 2nd ed. Cambridge: Cambridge Univ. Press.

European Commission. 2006. *Europeans and Their Languages*. Special Eurobarometer 243, Directorate General for Education and Culture, European Commission, http://ec.europa.eu/public_opinion/archives/ebs/ebs_243_en.pdf.

Grimes, J. E., and B. F. Grimes. 1993. *Ethnologue: Languages of the World*. 13th ed. Dallas: Summer Inst. Linguistics.

Hart-Gonzalez, Lucinda, and Stephanie Lindermann. 1993. "Expected Achievement in Speaking Proficiency, 1993." Manuscript, U.S. Dept. State, Foreign Service Inst., School Language Studies.

Hutchinson, William K. 2005. "Linguistic Distance as a Determinant of Bilateral Trade." *Southern Econ. J.* 72 (1): 1–15.
Van der Slik, Frans W. P. 2010. "Acquisition of Dutch as a Second Language: The Explanative Power of Cognate and Genetic Linguistic Measures for 11 West European First Languages." *Studies Second Language Acquisition* 32:401–32.