

A Dynamic Approach to the Determinants of Immigrants' Language Proficiency: The United States, 1980–2000¹

Frank van Tubergen
Utrecht University

Matthijs Kalmijn
Tilburg University

This article proposes a dynamic perspective on immigrants' language proficiency. Hypotheses are formulated about immigrants' language skills at arrival and about the speed with which immigrants learn the language thereafter. It pools data from the 1980, 1990, and 2000 U.S. Censuses, and uses a synthetic cohort design to analyze the language skills of immigrants within the first 20 years after migration. Multilevel models show that higher educated immigrants arrive with better language skills and learn the language quicker. Group size has a double-negative effect: it attracts less skilled immigrants, and it hampers language learning. These and other determinants are discussed in light of current research on immigrants' second-language proficiency.

INTRODUCTION

Studies have shown that many immigrants do not speak the language of the destination country well (Portes and Rumbaut, 2006). Language skills are important to study, because they play a pivotal role in immigrants'

¹We would like to thank the anonymous reviewers of *IMR* for valuable comments. Frank van Tubergen acknowledges financial support from the University of California Linguistic Minority Research Institute (UC LMRI) and the Arizona State University Linguistic Minority Research Initiative (ASU LMRI) of the Mary Lou Fulton College of Education. A previous version of this paper was presented at the conference "Immigrants, Education, and Language" on May 3–5, 2007, in Phoenix, Arizona, and at the Graduate School of Social Sciences, University of Bremen, Bremen (Germany), February 6, 2008.

economic incorporation (Shields and Price, 2002), and in ethnic intermarriage and interethnic contacts (Stevens and Swicegood, 1987). For these reasons, language proficiency is considered a fundamental aspect of immigrant integration (Alba and Nee, 2003; Bean and Stevens, 2003), and both sociologists and economists have frequently studied the determinants of immigrants' language proficiency (Esser, 2006; Chiswick and Miller, 2007).

Researchers have found that the language skills of immigrants depend on individual characteristics, such as age at migration and education, as well as contextual characteristics, like group size (Espenshade and Fu, 1997; Espinosa and Massey, 1997; Stevens, 1999). In this article, we aim to provide a better understanding of earlier findings by disentangling the language skills immigrants have when they arrive in the host country (what we label "*entry level*") from the changes in their language proficiency after migration (what we refer to as "*language acquisition*"). This generates a series of questions on the individual and contextual determinants that are suggested in previous research. For example, earlier studies found that higher educated immigrants and those living outside ethnic enclaves had better language skills. However, do higher educated immigrants have better language skills at the moment they arrive in the receiving country, do they learn the language more quickly after migration, or are both processes at work? Similarly, is the inverse relationship between relative group size and language skills an outcome of the sorting of immigrants with few language skills into "ethnic enclaves" or do immigrants in such communities learn the language less well?

Most previous research has used a static approach to language proficiency, in which immigrants' language proficiency is studied at a single point in time (*e.g.*, Dustmann, 1994; Chiswick and Miller, 1995; Espenshade and Fu, 1997; Espinosa and Massey, 1997; Stevens, 1999; Van Tubergen and Kalmijn, 2005). In such cross-sectional designs, it is impossible to distinguish immigrants' language proficiency at entry from their language acquisition after migration. In more recent times, a few studies have appeared that use panel data (Chiswick, Lee, and Miller, 2004; Hou and Beiser, 2006). Although these studies give valuable insights into the dynamics of language learning, they have their drawbacks too. The time span of the panel data is often rather short, ranging from less than three years (Chiswick, Lee, and Miller, 2004) to ten years (Hou and Beiser, 2006), which makes it difficult to study changes in immigrants' language skills over a longer time period. Furthermore,

sample sizes in panel studies are often quite small and restricted to specific ethnic groups (*e.g.*, South East Asian refugees in Hou and Beiser, 2006), making it difficult to examine contextual effects on the dynamics of language learning.

We use a synthetic cohort design to study language proficiency in a dynamic fashion. We pool multiple cross sections (*i.e.*, data from the 1980, 1990, and 2000 United States Censuses), which allows us to assess the language skills of immigrants at arrival and the changes in their skills thereafter. Because the cross sections we use cover a twenty-year period, we can separate effects of length of stay and effects of immigration cohort. Although the synthetic cohort design has been frequently used in studies on immigrants' economic incorporation (*e.g.*, Borjas, 1994), few studies have adopted this design to study language proficiency. One exception is Carliner (2000), who used 1980 and 1990 census data from the United States, but he did not study to what extent individual and contextual characteristics determined language skills at entry and language acquisition.

THEORY AND HYPOTHESES

Three General Mechanisms of Language Proficiency

Previous studies have proposed three general mechanisms to understand the impact of individual and contextual characteristics on the second language skills of immigrants (Stevens, 1999; Mesch, 2003; Esser, 2006; Chiswick and Miller, 2007). First of all, immigrants' second language skills are a direct function of the amount of *exposure* to that language. Thus, the more often immigrants are exposed to the host-country language, the better language skills they will have. For that reason, opportunities to hear, speak, read, and study the destination language are important for language learning (Stevens, 1992). Second, language proficiency is determined by economic *incentives*. Language is a form of human capital, with a clear positive effect on economic outcomes, but learning a new language is also costly. It is assumed that immigrants consider the costs of language learning and the expected benefits from such investments (Chiswick and Miller, 2001). Third, immigrants' language skills are an outcome of the *efficiency* by which immigrants learn that language. The more efficient immigrants are in learning a new language, the better they will speak that language given a certain amount of exposure.

Efficiency, according to researchers, is affected in turn by cognitive abilities but also by the difficulties of learning the language.

The mechanisms of exposure, incentives, and efficiency have been used to hypothesize about the role of individual and contextual determinants. In line with the theoretical model, it has been found that the language skills of immigrants are better among those who arrived at a young age, who are higher educated, and male. Researchers have also found that language skills are better when the language of the host country is an official language in the origin country, when the linguistic distance between the home and host languages is smaller, when the ethnic group in the region of residence is relatively small, and when people come from more economically developed, democratic, and globalized countries (*e.g.*, Dustmann, 1994; Chiswick and Miller, 1995; Hayfron, 2001; Dustmann and Fabbri, 2003; Hwang and Xi, 2008; Van Tubergen and Kalmijn, 2009).

We argue that immigrants' language proficiency is made up of two different components: language proficiency of immigrants at arrival in the host country (*entry level*), and the change in the language proficiency over time (*language acquisition*). The different mechanisms discussed above may have implications for both these components. We develop hypotheses on each determinant identified in prior research, specifying whether one would expect to find an effect on entry level (selection at arrival) and/or on language acquisition (learning after arrival).

Hypotheses on Individual Determinants

The positive effect of education on language proficiency can be interpreted in terms of pre-migration exposure, efficiency, and incentives (Esser, 2006). Higher educated people were more exposed to foreign languages before immigrating, they are better trained in learning languages, and they have more incentives to invest in language after migration because they generally enter non-manual jobs which require proficiency in the host language (Chiswick and Miller, 2001). The three mechanisms suggest that higher educated immigrants have better English-language skills because of their better command of the language when they arrive in the United States and, furthermore, that the higher educated learn the language more quickly after arrival.

The age at migration is primarily important for language learning. More specifically, it is assumed that immigrants who arrive at a younger age are more sensitive to learning new languages (Stevens, 1999).

Exposure after immigration may also play a role since young immigrants are more likely to attend school in the United States, which strongly affects language acquisition (Espenshade and Fu, 1997). As a result, we would expect that younger immigrants do not have higher language skills at entry, but they can be expected to learn the language more quickly.

Gender differences are usually interpreted in terms of incentives and exposure. In many immigrant groups, men have been more likely to migrate for economic reasons, whereas women have more often migrated for family reasons (Massey *et al.*, 1993). We would expect that men are better prepared for the labor market and thus have higher levels of language proficiency at entry. Furthermore, because men are more likely to participate in the labor market than women, they have more incentives to learn the language, and, as they are employed, are also more strongly exposed to the host language (Jasso and Rosenzweig, 1990; Espenshade and Fu, 1997). We expect that immigrant men arrive with better language skills and also learn the language at a faster rate than immigrant women, although we recognize that gender differences may vary across groups.

Hypotheses on Contextual Determinants

At the contextual level, we first look at the official status of English in immigrants' sending country (Van Tubergen and Kalmijn, 2005). In some countries, such as Australia and Canada, English is the official and dominant language. These countries are not included as immigrants from these countries all speak English. There are also countries in which English is official but not dominant (*e.g.*, India). For these origin groups, we expect that language skills will be better at arrival than for immigrants from countries in which English is neither official nor dominant. We furthermore expect that the language situation in the country of origin has no impact on language acquisition.

The positive effect of the degree of globalization of the country of origin has been interpreted in terms of prior exposure to foreign languages (Van Tubergen and Kalmijn, 2005). This argument is particularly relevant for immigrants in the United States, since English is the "world language," having the largest number of first- and second-language speakers (De Swaan, 2002). In many countries people are exposed to the English language through television, newspapers, books, education, and daily conversations. The amount of exposure, however, depends on the degree to

which a country is economically, politically, and culturally connected to the United States and other countries (Massey *et al.*, 1993). Some countries have a very open orientation, in which there are multiple and strong ties to other countries and their cultural and economic products, leading to more exposure to foreign languages and cultures. We argue that the degree of globalization of immigrants' country of origin has a positive effect on pre-migration exposure to English. Hence, globalization will affect the entry level but not the acquisition of further skills.

Relative group size is associated with post-migration opportunities to hear, speak, and study the second language (Stevens, 1992; Hwang and Xi, 2008). Group size directly affects the opportunities to communicate in the own ethnic language with neighbors, colleagues, and friends. It is also related to the availability of minority media, and to the opportunities of meeting and marrying a coethnic spouse (Stevens and Swicegood, 1987; Chiswick and Miller, 1996). Furthermore, group size is related to investments. Immigrants invest in the second language when such investments are expected to pay off economically in the (nearby) future. Such expectations are particularly low for immigrants who live and work in ethnic communities (Stevens, 1992). This implies that immigrants with fewer language skills are more likely to move to ethnically concentrated areas than immigrants who have better command of the English language (Bauer, Epstein, and Gang, 2005). For these reasons, we expect that a larger group will not only be detrimental for language acquisition, it will also attract less skilled immigrants and, hence, be associated with lower language levels at entry.

The role of linguistic distance between the official home and host language is mostly understood in terms of efficiency (Beenstock, Chiswick, and Repetto, 2001; Chiswick and Miller, 2005). It is argued that if the linguistic distance between two languages is small, the efficiency to learn the other language is higher. Hence, it should be more difficult for Chinese immigrants to learn English than for Italian immigrants, because Chinese is more distant from English than Italian. Thus, we assume that the more distant immigrants' official mother tongue is from English, the less those immigrants improve their English skills over time. Linguistic distance will have no effect on the entry level.

Some emigration flows are predominantly grounded on economic decisions, mostly followed by chain migration and family reunions. Other groups, typically referred to as "refugees," leave their country mainly because of war, discrimination, oppression, or other violations of political

rights and civil liberties. Researchers have argued that political migrants are less well prepared for participating in the labor market, and thus arrive with fewer language skills than those who migrated for economic reasons (Chiswick and Miller, 2001; Van Tubergen and Kalmijn, 2005). In addition, the unfavorable selection of political migrants and the traumatic events many of them have experienced may hamper their language investments. We expect that people from politically suppressive countries arrive with fewer language skills and also improve their English-speaking abilities less quickly over time than other migrants.

It has been argued that immigrants from poorer nations have received lower quality education, which is more difficult to transfer to the U.S. labor market. Transferability of educational diplomas plays a major role in ethnic disadvantage (Zeng and Xie, 2004). Furthermore, immigrants who have received lower quality education could be said to learn less well in general, and also to have more difficulties in participating in the U.S. school system than immigrants from wealthier nations. Thus, we expect that immigrants from poorer nations learn English slower after migration. These lines of reasoning are concerned with language acquisition, however, and not with language skills at entry.

DATA AND METHODS

Data

We use the 1980, 1990, and 2000 Census Public Use Samples, 5 percent files. The census data are well suited for our purposes. They contain immigrants from many origin countries, they allow us to study immigrants in different states, and they provide reasonably comparable questions on independent and dependent variables over time. Because the sample is large, we can focus on many different origin groups. The number of immigrant groups in our data is 156.

Dependent Variable

Speaking skills is measured in the census on a five-point scale. Respondents could report speaking only English at home, and if they don't, whether they speak English "very well," "well," "not well," or "not at all." Because this is not a continuous variable, we use a logistic regression model which distinguishes those who speak the language well, very well, or English only from those who speak the language not well or not at all.

This replicates the approach followed in previous studies (*e.g.*, Chiswick and Miller, 1992).

Length of Stay

Earlier research has shown that the largest increase in language learning occurs in the first few years after arrival, and that over time, the increase in language learning becomes smaller (Stevens, 1999). After experimenting with several forms of the duration-language link, Stevens (1999) concluded that taking the natural logarithm provides the most accurate representation. In this study, we use this specification as well. Because duration starts at 0 year, we first added 1 to the duration score before taking the log. A specific advantage of the logarithmic specification is that the coefficients can be interpreted in terms of percentages. Because the dependent variable is the log odds, the equation is a double-log equation (*e.g.*, Pindyck and Rubinfeld, 1991), which means that the coefficient for duration is the percentage increase in the odds of speaking the language well per percent change in length of stay.

Duration is calculated as the difference between the census year and the year of immigration. In the 1980 and 1990 censuses, immigration years were grouped so that no exact duration can be calculated. In 1980, 5-year categories were used and in 1990, 3-year categories were used. We assigned the category average to calculate duration scores. We also checked to see if the duration effect is sensitive to such categorizations. Results show that the duration effect is very similar in these two census years (1980 and 1990) compared to the most recent census where the year of immigration was recorded exactly. We select only those immigrants who have been in the United States for less than twenty years because after twenty years, little improvement in language proficiency occurs and the 1980 and 1990 censuses do not contain detailed categories of year of arrival for immigrants who arrived a long time ago.

Independent Variables

Years of Schooling. Measures the total years of completed schooling.

Age at Migration. Measures the age of the respondent at the time of immigration, in years.

Gender. Men are coded 1, women are coded 0.

English Language. To construct this variable, we use information on official language, *i.e.*, language used at school and in formal settings (Grimes, 2000) in the country of origin. Immigrants from dominant English-speaking countries are excluded. We distinguish two language situations: (1) countries in which English is an official language, and (2) countries where English is neither official nor dominant.

Globalization. We use part of the KOF index of globalization (Dreher, 2006), which intends to measure cultural globalization. It is composed of the number of McDonald's restaurants per capita (weighted by 40 percent), the number of Ikea per capita (40 percent), and the sum of exports and imports in books and pamphlets in percent of GDP (20 percent).

Relative Group Size. As an improvement to some earlier studies of group size, we constructed a variable that more directly measures the language environment. We measure the percentage of people with the same dominant, non-English language, based on the language situation in the country of origin.² This is measured at the state level, and for each census year, and refers to the population of 15 years and older.³ The language situation was obtained from Grimes (2000). Countries with no dominant language (*i.e.*, a language actively spoken by more than 40 percent of the population) were classified as 0. Inspection of the distribution of the variable revealed a U-shape pattern, with many smaller and bigger groups and relatively few groups in between these extremes. We therefore categorized the variable, classifying language groups smaller than 1 percent of the population (reference group), those between 1 and 5 percent of the population, and language groups larger than 5 percent.

Linguistic Similarity. To measure the difference from English with other languages, we build on a measure developed by Chiswick and Miller

²We prefer to use the dominant language in the country of origin rather than the language used at home to construct this measure, as has been done in several earlier studies (*e.g.*, Chiswick and Miller, 1992). The reason is that the languages people speak at home are endogenous to their language skills and the language skills of the people in their environment, whereas the official language in the home nation is not affected by these factors.

³The correlation between the measure we use (*i.e.*, percent same non-English language) and the measure used in most previous work (*i.e.*, percent own group) is 0.79.

(2005). The measure is based on empirically obtained information on how difficult it is for English speakers to learn foreign languages. This information was obtained in the context of language courses for students in the U.S. Chiswick and Miller (2005) report language scores for 43 languages of English-speaking Americans of average ability after 24 weeks. It ranges from 1 (lowest score) to 3 (highest score). Chiswick and Miller (2005) extended this list by matching languages not on the list to linguistically close languages for which scores are available (*e.g.*, Icelandic matched to Norwegian). We further extended their list, obtaining scores on all languages in our data set.

Political Suppression. A rating of political and civil rights (Karatnycky and Piano, 2002). Political rights varies from 1 (*e.g.*, free and fair elections, power for opposition) to 7 (*e.g.*, oppressive regime, civil war). Civil liberties varies from 1 (*e.g.*, freedom of expression and religion, free economic activity) to 7 (*e.g.*, no religious freedom, political terror, and no free association). We used the sum score for each country (2–14) for the 1980–1990 period.

Economic Development. We use the natural logarithm of Gross Domestic Product (GDP) per capita in constant U.S. dollars in 1980 as a measure of the economic situation in the origin country (World Bank, 2001).

We include controls for marital status (1 is married, 0 is divorced, widowed, single) and immigration cohort (ranging from 1954 to 2000, represented in 5-year categories). Table 1 presents descriptive figures of independent and dependent variables.

Methods

We use a synthetic cohort design to test our hypotheses. This design has frequently been used in studies on immigrants' economic assimilation. Two potential flaws of synthetic cohort designs need to be considered. First, there can be selectivity of immigration related to the outcome variable, in this case language skills. We deal with this issue by analyzing multiple censuses data, which allows us to control for immigration cohort. Furthermore, selectivity at arrival is also incorporated in our theoretical model and is the explicit focus of our study. Second, synthetic cohort designs do not follow the same individuals over time. This implies that we cannot address the problem of selective return migration and that we

TABLE 1
DESCRIPTIVE STATISTICS OF INDEPENDENT AND DEPENDENT VARIABLES

	Min	Max	Mean	SD
Dependent variables				
Not at all			0.14	
Not well			0.24	
Well			0.25	
Very well			0.30	
English only			0.07	
Independent variables				
Length of stay	0	20	9.59	5.78
Age at migration	0	90	26.46	13.66
Years of schooling	0	21	10.87	4.90
Male	0	1	0.51	
Globalization	1	85.68	26.10	12.44
GDP	442	22,319	6,143	3,505
Political suppression	2	14	7.93	3.15
Relative group size in state				
0–1%			0.40	
1–5%			0.18	
5–20%			0.42	
English official			0.15	
Language similarity	1	3	2.08	0.33
Control variables				
Married			0.61	
Immigration cohort				
1954–1965			0.03	
1966–1970			0.04	
1971–1975			0.10	
1976–1980			0.21	
1981–1985			0.18	
1986–1990			0.19	
1991–1995			0.13	
1996–2000			0.13	

Note: In the analysis, the variables of length of stay and GDP are logged, and all continuous variables are mean centered.

are unable to assess changes in the place of residence after immigration. However, research has shown that selective remigration hardly affects cross-sectional estimates of the determinants of immigrants' English skills in the U.S. (Lindstrom and Massey, 1994). To reduce possible bias due to selective migration within the U.S. after arrival, we analyze the size of immigrant groups at the state level instead of smaller geographical units that are more sensitive to selectivity.

Although we recognize the limitations of a synthetic cohort design, we note that panel data are still scarce at the moment. Moreover, panel data, have drawbacks too. First, panel designs rarely cover long periods of time. Our analysis is able to cover twenty years of experience in the host society, something which is in practice often impossible with panel data.

Second, panel designs are usually small, which makes it difficult to include a larger number of ethnic groups. Our analysis includes 156 groups and is therefore able to assess the impact of several contextual characteristics on language learning simultaneously. Panel designs are unable to address these effects and are therefore limited to individual-level and short-term effects on language learning.

We use multilevel (*i.e.*, hierarchical) modeling to estimate the relationships between language skills and the proposed individual and contextual characteristics. In most earlier studies, the role of contextual effects on language skills has been estimated without taking into account the multilevel structure of the data. Ignoring the clustering of observations leads to an underestimation of the standard errors, possibly resulting in incorrect conclusions about effects of contextual variables (Snijders and Bosker, 1999). We begin with a two-level multivariate logistic regression model, in which individuals i are nested within immigrant groups j . The dependent variable is θ_{ij} , which is the log odds of speaking the language well, *i.e.*, $\log(P_{ij} / 1 - P_{ij})$. The “traditional,” static approach to language proficiency can be formulated as follows:

$$\theta_{ij} = \beta_0 + \beta_l \ln L_{ij} + \beta_p \mathbf{X}_{pij} + \beta_q \mathbf{X}_{qij} + \beta_r \mathbf{X}_{rj} + u_{oj} + e_{oij}, \quad (1)$$

where θ_{ij} is the language proficiency of immigrants at the moment of the census, β_0 is the general mean, L_{ij} is the length of stay, \mathbf{X}_{pij} is a vector of level-one controls (year of migration and marital status), \mathbf{X}_{qij} are level-one predictors (age at migration, years of schooling, gender), and \mathbf{X}_{rj} are level-two predictors (political suppression, language of origin country, group size, globalization, GDP).⁴ The random part of the model consists of level 1 (e_{oij}) and level 2 (u_{oj}) residuals. Note that the year of entry is included in the control variables so that the effect of length of stay is not confounded by the effect of immigration cohort.

To test our hypotheses, we need to estimate effects of specified independent variables on 1) the level of language proficiency at entry, and 2) the change in language proficiency after entry. These two effects can be

⁴It should be noted that group size is not measured at level 2, because it varies within immigrant groups across states of residence. A cross-classified multilevel model with random intercepts for immigrant group and state of residence showed that the clustering within states is minimal (*i.e.*, intraclass correlation state = 0.0049). Hence, it is not necessary to include a random component for states.

estimated in a single model using interaction effects with length of stay. We estimate the following dynamic model:

$$\theta_{ij} = \beta_0 + \beta_l \ln L_{ij} + \beta_p \mathbf{X}_{p_{ij}} + \beta_q \mathbf{X}_{q_{ij}} + \beta_r \mathbf{X}_{r_{ij}} + \beta_{lq} (\ln L_{ij} \mathbf{X}_{q_{ij}}) + \beta_{lr} (\ln L_{ij} \mathbf{X}_{r_{ij}}) + u_{oj} + e_{oij}, \quad (2)$$

where $(\ln L_{ij} \mathbf{X}_{q_{ij}})$ and $(\ln L_{ij} \mathbf{X}_{r_{ij}})$ are the interactions between length of stay and the level 1 and level 2 explanatory variables, respectively. All level 1 and level 2 explanatory variables ($\mathbf{X}_{q_{ij}}$ and $\mathbf{X}_{r_{ij}}$) are centered around their mean. To see the dynamic implications of the model, we can rewrite equation (2) in terms of changes in language acquisition as follows:

$$\delta \theta_{ij} / \delta \ln L_{ij} = \beta_l + \beta_{lq} \mathbf{X}_{p_{ij}} + \beta_{lr} \mathbf{X}_{r_{ij}}. \quad (3)$$

Equation (3) yields a simple interpretation of the interaction effects:

1. β_l is the (logged) yearly change in language skills for the average respondent,
2. β_{lq} is the effect of level 1 predictors ($\mathbf{X}_{q_{ij}}$) on the change in language proficiency,
3. β_{lr} is the effect of level 2 predictors ($\mathbf{X}_{r_{ij}}$) on the change in language proficiency.

To evaluate the magnitude of the interaction effects, we calculate β_{lq} / β_l and β_{lr} / β_l , which indicates the degree to which the average length of stay effect is increased (or decreased) for each unit increase in $\mathbf{X}_{q_{ij}}$ and $\mathbf{X}_{r_{ij}}$, respectively (and controlling for other factors in the model).

The effects of these predictors on entry level language proficiency can be obtained by rewriting equation (2) for $\ln L_{ij} = 0$ (length of stay is zero):

$$\theta_{ij} | (\ln L_{ij} = 0) = \beta_0 + \beta_p \mathbf{X}_{p_{ij}} + \beta_q \mathbf{X}_{q_{ij}} + \beta_r \mathbf{X}_{r_{ij}} + u_{oj} + e_{oij}. \quad (4)$$

Equation (4) yields an interpretation of the main effects:

1. β_q is the effect of level 1 predictors ($\mathbf{X}_{q_{ij}}$) on the entry level language proficiency,

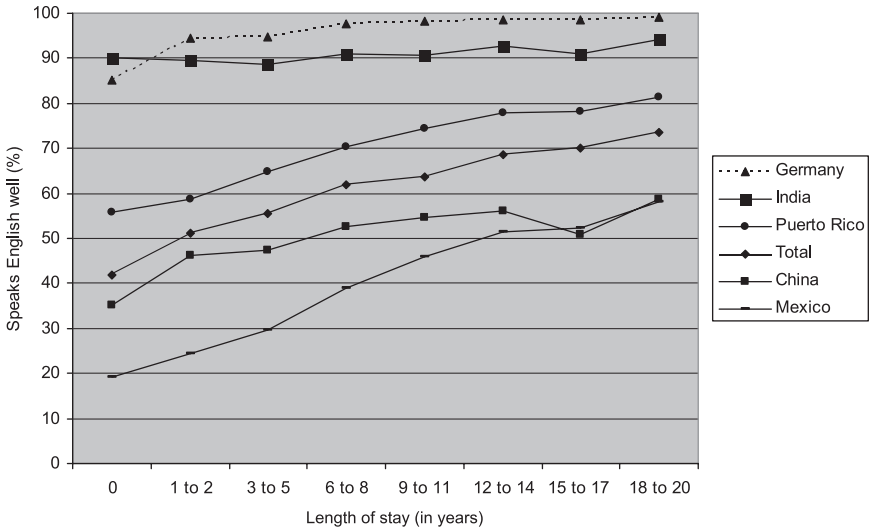
2. β_r is the effect of level 2 predictors (X_{ij}) on the entry level language proficiency.

We use the software program MLwiN 2.02 (Rashbash *et al.*, 2005) for our analysis. Results are obtained by first-order marginal quasi-likelihood (MQL) estimates (Snijders and Bosker, 1999).

RESULTS

Figure I presents the language skills of immigrants by years of residence (controlling for year of arrival). The results are imputations from a multivariate logistic model in which year of residence is represented with a set of dummy variables and in which year of arrival is included as a continuous variable (mean centered). Presented are the chances of speaking English (at least) well (*i.e.*, well, very well, or English only). The figure shows that of the entire immigrant population in the United States, slightly more than 40 percent speak English well upon arrival. There is a strong increase in language proficiency within the first two years, and a more steady increase thereafter. After 20 years in the U.S. about 75 percent speak English well.

Figure I. Language Proficiency of Immigrants from Non-English Speaking Countries by Length of Stay (Selected Groups). U.S. Census 1980–2000



Looking at the language proficiency of five important immigrant groups in more detail, we see strong group differences. Less than 20 percent of the Mexican immigrants speak English well at arrival, but after 20 years almost 60 percent speak English well. They end up with about the same language skills as Chinese immigrants, who arrived with better skills. Indians arrive with the best command of English, but their improvements over time are rather modest. Immigrants from Germany arrive with good language skills, quickly improve their language skills further, and virtually all speak English well after 20 years. These descriptive results show important group differences in both language skills at entry and language acquisition. To understand these differences, we need to look at the role of age at migration, education, group size, and other individual and contextual factors.

We present the results of two logistic multilevel regression models in Table 2. Model A reflects the "traditional approach" to language proficiency, estimating the effects of individual and contextual determinants on the *average* language proficiency (equation 1). The results of this model are compared to Model B, which is informed by our dynamic perspective. It includes the same determinants as in Model A, but adds the interactions with length of stay, allowing us to look at possible differential effects on *entry level* and on *language acquisition* (equation 2). For computational reasons we randomly selected 25 percent of the original 5 percent sample ($N = 1,560,276$), leading to 390,690 foreign-born immigrants of 18 years and older who were born in non-English-speaking countries. Additional analyses of smaller samples suggest that results are not sensitive to the number of cases.

To evaluate the magnitude of the interaction effects, we present the dynamic implications of Model B in Table 3. We first calculate how the length of stay effect is modified by a change in the independent variable. More specifically, we present the degree to which the average effect of duration (*i.e.*, the change in the odds of speaking well) is affected for each standard deviation increase in the independent variable. In other words, this number tells us how much faster or slower a person learns who is one standard deviation different in the independent variable. Moreover, we compare the effect of an independent variable on the entry level and on the ultimate level. More specifically, we present the relative change in the log odds of speaking well for each standard deviation change in the independent variable for a duration of 0 (at entry) and for a duration of 20 years.

TABLE 2
 MULTIVARIATE MULTILEVEL LOGISTIC REGRESSION OF IMMIGRANTS' LANGUAGE SKILLS: SPEAKING (VERY)
 WELL OR FLUENT VERSUS POORLY OR NOT AT ALL

	Model A		Model B			
	<i>b</i>	SE	Main Effect		Interaction with L	
			<i>b</i>	SE	<i>b</i>	SE
Individual variables						
Length of stay (ln)	0.511*	0.008	0.588*	0.015		
Age at migration	-0.040*	0.000	-0.003*	0.001	-0.017*	0.000
Years of schooling	0.135*	0.001	0.082*	0.003	0.024*	0.001
Gender (male = 1, female = 0)	0.208*	0.008	0.062*	0.025	0.069*	0.012
Contextual variables						
Globalization	0.015*	0.005	0.011*	0.005	0.003*	0.001
GDP (ln)	-0.095	0.118	-0.172	0.121	0.040*	0.013
Political suppression	-0.014	0.022	-0.048*	0.023	0.017*	0.002
Relative group size (versus 0–1%)						
1–5%	-0.274*	0.017	-0.191*	0.040	-0.042*	0.018
5–20%	-0.358*	0.020	-0.088*	0.039	-0.136*	0.017
English official	0.859*	0.176	0.985*	0.182	-0.082*	0.024
Language similarity	0.377*	0.169	0.171	0.175	0.109*	0.023
Control variables						
Married	-0.090*	0.009	-0.115*	0.009		
Cohort (1954–1965 is reference)						
Immigrated 1966–1970	-0.155*	0.037	-0.157*	0.038		
Immigrated 1971–1975	-0.076*	0.033	-0.079*	0.034		
Immigrated 1976–1980	-0.187*	0.032	-0.200*	0.033		
Immigrated 1981–1985	-0.267*	0.032	-0.280*	0.033		
Immigrated 1986–1990	-0.282*	0.032	-0.296*	0.033		
Immigrated 1991–1995	-0.323*	0.033	-0.329*	0.034		
Immigrated 1996–2000	-0.175*	0.035	-0.184*	0.036		
Intercept	0.561*	0.086	0.448*	0.090		
Variance of intercept	0.521	0.073	0.533	0.075		
N _j (group)	156		156			
N _i (individual)	390,690		390,690			

Note: **p* < 0.05 (two-tailed test).

Individual Determinants

Our results of the traditional approach to immigrants' second-language proficiency (Model A) replicate a well-known pattern: age at migration has a negative effect on the average language proficiency. One standard deviation increase in the age at arrival (*i.e.*, 13.66, Table 1) is associated with a 42 percent decline in the odds of speaking the language (at least) well [calculated as $1 - \exp(-0.040 \times 13.66)$].

We expected that the negative effect of age at migration results from language differences after migration, not from entry differences. To examine this hypothesis, we look at Model B. As expected, the dynamic

TABLE 3
DYNAMIC IMPLICATIONS OF MULTIVARIATE MULTILEVEL LOGISTIC REGRESSION OF IMMIGRANTS'
LANGUAGE SKILLS

	Effect on Log Odds Per SD Change in X		Change in the Duration Effect Per SD Change in X
	At Entry	After 20 years	
Age at migration	-0.04	-0.74	-0.45
Years of schooling	0.40	0.75	0.23
Gender (male = 1, female = 0)	0.06	0.27	0.14
Globalization	0.14	0.25	0.07
GDP (ln)	-0.11	-0.03	0.05
Political suppression	-0.15	0.01	0.10
Relative group size (versus 0-1%)			
1-5%	-0.19	-0.32	-0.08
5-20%	-0.09	-0.50	-0.27
English official	0.99	0.74	-0.16
Language similarity	0.06	0.16	0.07

Notes: The average duration effect is obtained from Model A, Table 2. Other numbers from Model B.

specification reveals that age at migration hardly affects the language skills of immigrants at arrival. The main effect of age in the interaction model is very small. Furthermore, and equally in line with our expectations, there is a significant and negative interaction between length of stay and age at arrival. People who arrive at an older age learn the new language less quickly than those who arrive at a younger age. To evaluate the magnitude of this effect, we calculate how the length of stay effect is modified by the age at arrival. For an average immigrant, the increase in language proficiency over time is $b = 0.51$ (Model A, Table 2). For each standard deviation increase in the age at arrival, this increase effect is reduced by 45 percent [calculated as $(-0.017 \times 13.66) / 0.51$]. This shows that language acquisition is strongly affected by age at arrival. We can also compare the effect on the entry level with the effect on the ultimate level. One standard deviation increase in the arrival age is associated with a decline of 0.04 in the log odds of speaking the language well at arrival and this increases to a decline of 0.74 in the log odds after 20 years (Table 3).

With respect to years of schooling, we find a positive relationship (Table 2, Model A), showing that people who have received more education generally have more command of the language. In standardized terms, an increase of one standard deviation in education leads to a doubling of the odds to speak the language well [$\exp(0.135 \times 4.9) = 1.94$]. We expected that this relationship results from two forces: higher educated immigrants arrive with better language skills and they learn the

language faster than lower educated immigrants. The results of Model B support this. The main effect of education is significant, implying that the higher educated arrive with better language skills. The interaction effect is significant as well and positive, showing that the higher educated learn more quickly. The size of the interaction effect is not trivial. A one standard deviation increase in schooling leads to a 23 percent increase in the effect of length of stay (Table 3). Moreover, at entry, a one standard deviation difference in education is associated with a 0.40 higher log odds of speaking the language and this increases to a 0.75 higher log odds after 20 years of stay (Table 3).

The effect of gender is significant and strong: Men have a 23 percent higher odds to speak English well than women. The language disadvantage of immigrant women was observed in earlier work, and we expected that women would arrive with fewer skills and learn the language less quickly than men. Model B indeed shows that gender differences already exist at entry. We also see, as expected, a significant effect on the change over time. Table 3 shows that the duration effect is 14 percent stronger for men than for women. This is in line with the hypothesis.

Contextual Determinants

We now turn to the contextual determinants. To begin, Model A shows that, as observed in earlier research, globalization positively and significantly affects the language skills of immigrants. The more globalized the immigrants' country of origin, the better their language skills in the United States. A one standard deviation increase in globalization is associated with a 21 percent increase in the odds of speaking the language well. We expected that globalization only has an effect on entry differences in language skills, not on language acquisition. Our dynamic specification of language (Model B) indeed shows that globalization has a positive and significant impact on entry level differences. More specifically, a one standard deviation increase in globalization is associated with a .14 increase in the log odds at entry (Table 3). Furthermore, we find, as expected, that globalization has no effect on language learning once immigrants are in the United States. The interaction effect of globalization and length of stay in Table 2 is significant but small.

Contrary to some earlier studies, we find limited evidence for the role of economic development in the country of origin. In Model A, GDP per capita has no significant effect on the language skills of

immigrants. Model B shows a small positive effect of GDP on language acquisition. Earlier studies conducted in the United States reported a more positive impact of economic development on immigrants' language proficiency (Jasso and Rosenzweig, 1990; Espenshade and Fu, 1997). One explanation is that we control for the clustering of observations at the immigrant group level. When we estimate a simple binary logistic model without a multilevel design, we replicate the earlier observed positive impact of GDP per capita. Hence, we conclude that, net of other factors in our model, economic development has little effect on group differences in language proficiency.

We find that political suppression in the country of origin has no overall effect on language fluency (Model A). It appears that initial differences in language are fully offset by opposite forces after arrival. Model B shows that political suppression has a significantly *negative* effect at language skills at entry, and a significantly *positive* interaction effect with duration. In other words, immigrants from politically suppressive countries enter with poorer language skills as expected but they learn more quickly. A one standard deviation increase in the level of suppression is associated with a 0.15 lower *initial* log odds of speaking the language well. After 20 years, the effect of political suppression has declined to a change in the log odds of 0.01. Hence, immigrants from politically suppressed origin countries catch up during their stay in the United States.

We see that group size has a significant negative effect on language proficiency. The higher the percentage of people who have the same non-English mother tongue within the state of residence, the less well immigrants speak English. Immigrants who live in a state with 1–5 percent of the population who speak the same non-English language have a 24 percent lower odds to speak English well than those who live in a state with less than 1 percent same-language speakers. For those above 5 percent same-language speakers, the odds are 30 percent lower. Our hypothesis was that this well-known empirical regularity is the result of both entry and acquisition processes: immigrants with lower language skills would be attracted to non-English areas, and in these areas, immigrants would learn the English language less well. Model B supports our hypothesis.⁵ We see

⁵In a separate analysis, we checked whether the same patterns were found when using absolute group size. This is indeed the case: the absolute number of people who speak the same non-English language in a state has a negative effect on language skills at entry and language acquisition.

a significant negative effect on entry level and a significant and negative interaction effect with length of stay.⁶

We also find important differences between the language situations in the country of origin. It appears that immigrants who come from countries in which English is official (but not the dominant language) have a 2.4 times higher odds to speak the language well than immigrants who come from countries in which a language other than English is dominant and official. We predicted that differences across language origins come about through entry differences, not through language acquisition. Indeed, Model B shows important differences in entry level. Immigrants who come from countries in which English is an official language have a 2.7 times higher odds to speak English at arrival. In addition, we see a significantly negative interaction with length of stay. This means that the increase in language learning among immigrants from officially English countries is smaller than among immigrants from countries in which English is not an official language. This unexpected finding may reflect a ceiling effect: immigrants from officially English countries speak the language already quite well at arrival and cannot improve their language skills much more over time. Hence, these immigrants gradually lose their advantage to those who arrived with fewer skills, although after 20 years there is still a substantial difference (Table 3).

Finally, we examine the role of linguistic distance. Model A shows that, as observed in previous research, immigrants whose native language is more similar to English have a higher odds to speak English well. One standard deviation increase in language similarity is associated with a 13 percent higher odds of speaking English well. From a dynamic perspective, we expected that linguistic distance would lower the speed with which immigrants learn a new language, but that distance would not affect entry differences. In line with this hypothesis, we see a positive and significant interaction effect in Model B. Hence, immigrants from the linguistically most similar countries have no advantage at the moment they arrive in the U.S. They do, however, learn English somewhat quicker.

⁶The better language skills at arrival of immigrants who belong to the largest groups (*i.e.*, 5–20 percent) in comparison with immigrants from intermediate groups (*i.e.*, 1–5 percent) can be explained by unmeasured pre-migration exposure to English among Mexicans, who belong to the largest group. When we leave out Mexicans from our sample, we see that the disadvantage at entry for the intermediate groups (main effect: $b = -0.252$) is almost the same as for the largest groups ($b = -0.243$).

DISCUSSION AND CONCLUSION

Previous work has documented the importance of several individual and contextual factors in explaining immigrants' skills in the language of the host society. Most prior studies, however, have used a static approach to immigrants' language proficiency, predicting the language skills at a certain point in time, without considering the dynamics of second language learning. In this study, we pooled 1980, 1990, and 2000 census data from the United States. Using a synthetic cohort design while controlling for immigration cohort, we examined the language skills of immigrants at the moment of arrival in the United States and their improvement over a twenty-year period. In this way, we gain a better understanding of what determines language proficiency at *entry* (selection at arrival) and the processes that are at work in *language acquisition* once immigrants are in the host society. This provides a new perspective on the determinants of language proficiency. For example, do higher educated immigrants speak English already well when they arrive in the U.S., or do they learn English much faster than lower educated immigrants?

Our analyses show that immigrants strongly differ in their command of the English language when they arrive in the United States, while they also strongly differ in the speed with which they learn English after migration. Using the more general theoretical notions on exposure, economic incentives, and efficiency, several trajectories can be identified.

To begin, there are groups that speak the language already well upon arrival and hardly improve their English skills any further. This is obviously the case for immigrants from dominant English countries, who are excluded from our study, but also, to a lesser extent, for immigrants from officially English countries. Immigrants from these countries were more exposed to English prior to migration and still have a language advantage after 20 years compared to immigrants from non-officially English countries.

There are also groups that show no difference at the moment of arrival, but once in the U.S., they differ in their language learning. We find that age at migration and linguistic distance have no impact on language differences at arrival, but they do affect language acquisition. Thus, immigrants who arrived at a young or old age do not differ in their language proficiency at arrival. However, immigrants who migrated at a young age learn English much faster in the United States. Younger immigrants are more efficient in language learning and they are more exposed to English after migration (*e.g.*, at school in the U.S.). A similar, though less strong,

pattern is observed for linguistic distance. There are no differences between immigrants with a mother tongue very different or very close to English at arrival. However, immigrants who have a mother tongue very different from English, such as the Chinese, have more difficulties in learning English and this hampers their language learning in the U.S.

Some groups have a language disadvantage at arrival, but completely catch up during their stay in the United States. Our study shows that political migrants speak English less well when they arrive, but within 20 years speak English as good as other groups. Political migrants are less well prepared when they arrive in the United States, leading to fewer skills at arrival. That political migrants learn English faster is surprising, as many of them have experienced traumas. Possibly, however, political migrants have stronger intentions to stay than other immigrants, making it more attractive for them to invest in language learning.

Finally, we see several groups for which entry differences in language proficiency become more pronounced over time. Such “double disadvantages” are observed for immigrants in larger groups, for women, and for the lower educated. Larger immigrant groups attract immigrants with fewer language skills, but group size also negatively affects language learning. Immigrants who live in a state with a larger proportion of immigrants speaking the same non-English mother tongue are less often exposed to English, and they have fewer incentives to invest in the English language, as they can use their first language in the neighborhood, at work, and with friends.

Immigrant women have a “double disadvantage” as well: they arrive with fewer English skills, and they learn English less quickly as immigrant men — presumably because of economic incentives. Similarly, we see that lower educated immigrants arrive with fewer language skills and learn English less quickly than higher educated immigrants. Lower educated people are less efficient in learning new languages, less exposed to languages, and can find jobs for which only limited knowledge of English is required.

In closing, we would like to argue that the distinction we have made between language proficiency upon arrival and language assimilation later on is of more general importance and calls for new lines of research on immigrant integration. The distinction between differences at entry and differences in assimilation can be applied to other aspects of language proficiency, such as reading and writing skills, and to patterns of language use (Akresh, 2007). Furthermore, this dynamic perspective can be generalized to others dimensions of sociocultural integration, such as political participation, norms and values, and interethnic contacts.

REFERENCES

- Akresh, I. L.
2007 "Contexts of English Language Use among Immigrants to the United States." *International Migration Review* 41:930–955.
- Alba, R., and V. Nee
2003 *Remaking the American Mainstream: Assimilation and Contemporary Immigration*. Cambridge, MA: Harvard University Press.
- Bauer, T., G. S. Epstein, and I. N. Gang
2005 "Enclaves, Language, and the Location Choice of Migrants." *Journal of Population Economics* 18:649–662.
- Bean, F., and G. Stevens
2003 *America's Newcomers and the Dynamics of Diversity*. New York: Russell Sage Foundation.
- Beenstock, M., B. R. Chiswick, and G. L. Repetto
2001 "The Effect of Linguistic Distance and Country of Origin on Immigrant Language Skills: Application to Israel." *International Migration* 39:33–60.
- Borjas, G. J.
1994 "The Economics of Immigration." *Journal of Economic Literature* 32:1667–1717.
- Carliner, G.
2000 "The Language Ability of U.S. Immigrants: Assimilation and Cohort Effects." *International Migration Review* 34:158–182.
- Chiswick, B. R., Y. L. Lee, and P. W. Miller
2004 "Immigrants' Language Skills: The Australian Experience in a Longitudinal Survey." *International Migration Review* 38:611–654.
- , and P. W. Miller
2007 *The Economics of Language: International Analyses*. London: Routledge.
- , and ———
2005 "Linguistic Distance: A Quantitative Measure of the Distance Between English and Other Languages." *Journal of Multilingual and Multicultural Development* 26:1–11.
- , and ———
2001 "A Model of Destination-Language Proficiency Acquisition: Application to Male Immigrants in Canada." *Demography* 38:391–409.
- , and ———
1996 "Ethnic Networks and Language Proficiency among Immigrants." *Journal of Population Studies* 9:19–35.
- , and ———
1995 "The Endogeneity between Language and Earnings: International Analyses." *Journal of Labor Economics* 13:246–288.
- , and ———
1992 "Language in the Immigrant Labor Market." In *Immigration, Language, and Ethnicity: Canada and the United States*. Ed. B. R. Chiswick. Washington, DC: American Enterprise Institute. Pp. 229–296.

- De Swaan, A.
2002 *The World Language System: A Political Sociology and Political Economy of Language*. Cambridge: Polity Press.
- Dreher, A.
2006 "Does Globalization Affect Growth? Empirical Evidence from a New Index." *Applied Economics* 38:1091–1110.
- Dustmann, C.
1994 "Speaking Fluency, Writing Fluency and Earnings of Migrants." *Journal of Population Economics* 7:133–156.
———, and F. Fabbri
2003 "Language Proficiency and Labour Market Performance of Immigrants in the UK." *The Economic Journal* 113:695–717.
- Espenshade, T. J., and H. Fu
1997 "An Analysis of English-Language Proficiency among U.S. Immigrants." *American Sociological Review* 62:288–305.
- Espinosa, K. E., and D. S. Massey
1997 "Determinants of English Proficiency among Mexican Migrants to the United States." *International Migration Review* 31:28–50.
- Esser, H.
2006 *Sprache und Integration: Die sozialen Bedingungen und Folgen des Spracherwerbs von Migranten*. Frankfurt/New York: Campus Verlag.
- Grimes, B. F.
2000 *Ethnologue: Languages of the World*, 14th ed. Dallas: SIL International.
- Hayfron, J. E.
2001 "Language Training, Language Proficiency and Earnings of Immigrants in Norway." *Applied Economics* 33:1971–1979.
- Hou, F., and M. Beiser
2006 "Learning the Language of a New Country: A Ten-year Study of English Acquisition by South-East Asian Refugees in Canada." *International Migration* 44:135–165.
- Hwang, S.-S., and J. Xi
2008 "Structural and Individual Covariates of English Language Proficiency." *Social Forces* 86:1079–1104.
- Jasso, G., and M. R. Rosenzweig
1990 *The New Chosen People: Immigrants in the United States*. New York: Russell Sage Foundation.
- Karatnycky, A., and A. Piano
2002 *Freedom in the World: The Annual Survey of Political Rights and Civil Liberties, 2001–2002*. Piscataway, NJ: Transaction Publishers.
- Lindstrom, D. P., and D. S. Massey
1994 "Selective Emigration, Cohort Quality, and Models of Immigrant Assimilation." *Social Science Research* 23:315–349.
- Massey, D. S. *et al.*
1993 "Theories of International Migration: A Review and Appraisal." *Population and Development Review* 19:431–466.

- Mesch, G. S.
2003 "Language Proficiency among New Immigrants: The Role of Human Capital and Societal Conditions." *Sociological Perspectives* 46:41–58.
- Pindyck, R. S., and D. L. Rubinfeld
1991 *Econometric Models and Economic Forecasts*. New York: McGraw-Hill.
- Portes, A., and R. G. Rumbaut
2006 *Immigrant America: A Portrait*, 3rd ed. Berkeley: University of California Press.
- Rashbash, J., F. Steele, W. Browne and B. Prosser
2005 *A User's Guide to MLwiN Version 2.0*. Bristol: Center for Multilevel Modelling.
- Shields, M. A., and S. W. Price
2002 "The English Language Fluency and Occupational Success of Ethnic Minority Immigrant Men Living in English Metropolitan Areas." *Journal of Population Economics* 15:137–160.
- Snijders, T., and R. Bosker
1999 *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. London: Sage Publications.
- Stevens, G.
1999 "Age at Immigration and Second Language Proficiency among Foreign-born Adults." *Language in Society* 28:555–578.
- 1992 "The Social and Demographic Context of Language Use in the United States." *American Sociological Review* 57:171–185.
- , and G. Swicegood
1987 "The Linguistic Context of Ethnic Endogamy." *American Sociological Review* 52:73–82.
- Van Tubergen, F., and M. Kalmijn
2005 "Destination-Language Proficiency in Cross-National Perspective: A Study of Immigrant Groups in Nine Western Countries." *American Journal of Sociology* 110:1412–1457.
- , and ———
2009 "Language Use and Proficiency of Immigrants in the Netherlands: Opportunities or Incentives?" *European Sociological Review* 25:169–182.
- World Bank
2001 *World Development Indicators*. Washington, DC: CD-ROM World Bank.
- Zeng, Z., and Y. Xie
2004 "Asian-Americans' Earnings Disadvantage Reexamined: The Role of Place of Education." *American Journal of Sociology* 109:1075–1108.