

**National Center for Border Security and Immigration**  
**Research Lead: The University of Arizona (Tucson, Arizona)**



**BORDERS Awards in Immigration Research:  
New Immigrant Survey Final Report**

---

*Assessing the Influence of Language Background on Immigrants' English  
Language Proficiency*

---

Aaron Ponce  
Indiana University  
December 31, 2012

**Department of Homeland Security – Grant No. 2008-ST-061-BS0002**

*"This research was supported by the United States Department of Homeland Security through the National Center for Border Security and Immigration under grant number 2008-ST-061-BS0002. However, any opinions, findings, and conclusions or recommendations in this document are those of the authors and do not necessarily reflect views of the United States Department of Homeland Security."*

## **TABLE OF CONTENTS**

Abstract	3
Introduction	3
Results/Findings	8
Limitations/challenges	15
Policy recommendations	15
Next steps in research	16
Works cited	17

## Abstract

Immigrants' integration into a host society's labor market and civil society is fundamentally dependent on the acquisition of vital language skills. New Immigrant Survey 2003 restricted data are used to examine the factors affecting adult immigrants' English language proficiency. The study examines how immigrants' own language backgrounds affect their English proficiency along two different dimensions: speaking and understanding English. In particular, a measure of the sound-based *phonemic diversity* of various immigrant languages is used to examine how immigrants' childhood languages and current home languages affect their levels of English proficiency while controlling for typical sociodemographic differences. Results show that immigrants who speak languages with higher phonemic diversity have higher levels of English proficiency on average.

## Introduction

Immigrants to the developed world face a number of challenges to successful integration into host societies (Alba and Nee, 2003; Bean and Stevens, 2003; Jasso and Rosenzweig, 1990; Portes and Rumbaut, 2006). Among the most significant and immediate challenges is acquiring the necessary language skills in the host language to fully participate in economic, social, and political life. Social scientists have noted the difficulties acquiring language proficiency as both an outcome associated with and a determinant of successful incorporation.

Early work on language proficiency focuses on perhaps the most material aspect of incorporation, labor market integration. Research shows that immigrants with lower levels of English proficiency typically earn less in wages than their English-proficient counterparts, even after accounting for skill differentials and nativity (Grenier, 1984a; McManus et al., 1983; McManus, 1985; Tainer, 1988). More generally, lack of English language ability has been linked to lower levels of occupational prestige and upward mobility (Akresh, 2008; Dávila and Mora, 2000; Kossoudji, 1988). This means that many immigrants coming to the U.S. are at a comparative disadvantage when competing in the labor market due to their language proficiency status. Outside of the American context, similar findings have been confirmed in Germany (Dustmann, 1994; Dustmann and van Soest, 2002), Australia (Chiswick and Miller, 1995), Canada (Chiswick and Miller, 2007a), and the United Kingdom (Shields and Price, 2002).

The effects associated with limited English proficiency extend well beyond the economic realm, however. In addition to labor market outcomes, an immigrant's level of English language proficiency is found to be tied to other important indicators of well being and health (Akresh, 2007a; Arcia et al., 2001). Language proficiency is also tied to levels of naturalization and political participation (Bloemraad, 2002, 2006; Cho, 1999). Such negative effects may even carry over to generations beyond the first generation. That is, these detrimental consequences may affect family members and present persistent disadvantages for the second generation as well (Bleakley and Chin, 2004, 2008; Casey and Dustmann, 2008). Evidence overwhelmingly points to the importance of language in immigrants' post-migration lives. Indeed, this makes sense considering that very little of the social experience is removed from language-based, communicative processes.

Naturally, additional questions emerge: *Who* among immigrants is more likely to have higher levels of English proficiency, and what exactly affects these levels? Though perspectives and approaches differ somewhat, a number of robust findings emerge regarding the sociodemographic indicators of English proficiency among immigrants. On average, immigrants with higher levels of English proficiency include: those who have resided in the U.S. longer, those who arrived at a younger age, males, those with higher levels of education, those married to a native speaker of English or with certain household compositions (with children in some instances), and those who reside in geographic areas with lower concentrations of co-ethnics (Chiswick and Miller 1998b; Espenshade and Fu 1997; Espinosa and Massey 1997; Grenier 1984a; Lindstrom and Massey 1994; Stevens 1992, 1999b). While economic studies emphasize language proficiency as key to the more perfect transferability of skills and human capital, sociological explanations tend to focus on structural impediments to language exposure and the opportunities to learn a host language. In both accounts, education level, age at migration, and years since migration in particular display strong effects on English language proficiency among immigrants. This suggests that exposure not only to the host society, but perhaps also to pre-migration environments that reward English language use or processes of acculturation through education, make a difference.

Although both lines of research provide a wealth of information revealing the social groups most likely to learn English, less can be said about how immigrants from different language groups achieve English proficiency along different linguistic trajectories. Fortunately, important contributions by Chiswick and colleagues have attempted to remedy this oversight using new concepts. In particular, this research has established that the linguistic distance between immigrants' native languages and English—i.e., “the extent to which languages differ from each other”—a significant, negative effect on English language proficiency (Chiswick and Miller 1998a, 2005, 2007b, 2008). These results hold net of other individual sociodemographic factors such as education, years since migration, and age at arrival. For instance, Mandarin Chinese speakers who speak a language that is relatively distant from English (linguistic distance of 0.67) are less likely to become proficient in English than, say, a speaker of French (linguistic distance of 0.40) (scores from Chiswick and Miller 2007a, p. 579). Linguistic distance is presumed to work through the relative difficulty of learning a language, with difficulty suppressing levels of English proficiency. Such a conceptualization, however, comes with a set of assumptions that may not withstand sustained scrutiny. The contributions by Chiswick and colleagues nevertheless motivate the present study by demonstrating that it is not simply the characteristics that individuals bring net of their linguistic backgrounds, but linguistic background itself that is associated with language proficiency. This study extends these findings to show that there are important structural properties of language that can directly be measured to test the relationship between language background and English proficiency.

While the linguistic distance measure represents a watershed in the study of language background and proficiency, additional work must be done to determine the actual linguistic properties inherent in immigrants' native languages that are related to their eventual proficiency in a non-native language. For this, a universal measure of well-defined linguistic structures is necessary. Research in linguistics can be used as a guide for formulating effective and useful measures of language differences. Linguists' study of language relatedness by families provides a starting point for understanding how languages differ. Historical linguists, for example, trace relatedness between any pair of languages by comparing individual words from a set list in the two languages to quantify their structural similarities. The presumption here is that if there are a significant number of structural similarities between the words from the two languages' word lists, then the languages share a common historic origin. The degree of relatedness between pairs of languages is then systematically analyzed and mapped onto language families (for a detailed description, see Bakker et al. 2009). Such methods of measuring the relatedness of languages have often been used to trace languages

back in time to particular geographic areas of the world, often referred to as language homelands (Wichmann et al. 2010). For the purposes of the current study, this literature shows that the worlds' languages are divided into families that provide a general yet important understanding of how different or related such languages are. While the standard *linguistic distance* measure used in the immigration literature is often touted as an improvement over the categorization of language families (Chiswick and Miller 2001, 2007a), there is reason to believe that language families may also be important. What is at question, then, is whether the linguistic distance measure captures something other than relatedness by language family. Including both the linguistic distance measure and language families in the statistical analyses allows me to test whether this is true.

Since the outcome of interest in the current study is English proficiency among non-native speakers, I also look to the linguistics literature on second language acquisition. Studies in second language acquisition suggest that *language transfer* affects the learning of a non-native language. Briefly defined, language transfer occurs when linguistic structures from a person's native language are carried over to the non-native language that an individual is learning (Gilbert 1983; Odlin 1989; Romaine 2003). Although language is complex and encompasses a number of different acoustic, cognitive, and social dynamics, linguists have made substantial headway on the imposing task of breaking down the complexity of language into relatively simple components to study. In determining which domain of language to examine, there is evidence that *phonetics* and *phonology* are linguistic systems that are especially prone to language transfer (Dulay et al. 1982). In basic terms, *phonetics* and *phonology* are the areas of language involving the different acoustic sounds that a language contains, and how these sounds are structured and related to one another. I focus on this area of language for the current study and use a measure of *phonemic diversity* to gauge the structural phonological differences between immigrants' native languages. Phonemic diversity captures the sound-based complexity of a language through a language's inventory of meaningful sounds—consonants, vowels, and tones. It thus allows for an analysis of how the phonological structure of an immigrants' native language may carry over to the learning of a non-native language, in this case English.

### Methodology

For the analyses, survey data are used from the 2003 New Immigrant Survey (NIS). In particular, the analyses use NIS Restricted Data version 1, which includes an impressive array of data reflecting the many diverse languages that immigrants to the U.S. speak. In all, over 200 languages are represented in the original sample. The sample includes new immigrants to the United States aged 18 and older who were granted legal permanent resident status between May and November 2003. Interviews were conducted in the language of the respondent's choice, and encompassed some 95 interview languages (Jasso 2011). The original full sample of adult respondents includes 8573 new immigrants to the U.S. with a response rate of 69%.

I delimit the original sample in stages for the current analysis. First, I drop respondents who were interviewed overseas from all analyses since they did not receive the full battery of language-related items. Second, I drop respondents who are considered monolingual English speakers from all analyses. This includes those who answered "no" to the question "Have you ever spoken any language other than English?" Such respondents represent immigrants from predominantly English-speaking countries. Additionally, I use the two main explanatory variables to delimit the sample further by language background. If respondents reported speaking English when they were 10 years old (childhood language) *and* reported speaking English currently in the home (home language), this is taken as an indication that a non-English language does not figure prominently in the respondent's level of English proficiency. Accordingly, I exclude these respondents from all analyses.

In total, observations dropped due to missing data on the main outcome and explanatory variables or covariates results in a reduction of the sample by 23% for models assessing the relationship between childhood language and speaking/understanding English (N=6607). For models assessing the relationship between home language and speaking/understanding English, there is a 36% reduction of the sample (N=5521).

The main explanatory measure used, phonemic diversity, measures each respondent's reported childhood language and home language in terms of how complex the phonological structures of these languages are. In addition to the phonemic diversity measure, I include the established linguistic distance measure in specified models. Because the linguistic distance measure is more limited in the number of languages it measures, fewer languages could be assigned a linguistic distance value resulting in smaller samples for models testing its effect.

The outcome variables are self-reported proficiency in speaking English and understanding English. Each respondent was asked how well s/he spoke and understood English with responses ranging from (1) not at all, (2) not well, (3) well, to (4) very well. For the main models presented, respondents answering either "well" or "very well" are assigned a value of 1, while respondents answering "not at all" or "not well" are assigned a 0.

The main explanatory variables examined are the phonemic diversity of the respondent's childhood language and home language. Childhood language is the language that respondents report speaking when they were 10 years old. Home language is the language that respondents report currently speaking in the home. Examining these variables allows for an analysis of language background effects by period and current relevance; that is, by whether language effects can be traced back to early childhood and/or whether exposure to and use of the language are recent. This extends previous studies of English proficiency among immigrants, since prior research using different datasets has generally lacked adequate data to investigate the multiple facets of language background.

Phonemic diversity data come from the World Atlas of Language Structures (WALS) (Dryer and Haspelmath 2011). The WALS project quantifies and maps the structural properties of the world's languages through expert assessments of descriptive material in linguistics. WALS data have been used for key studies in the origins and dissemination of human language across the globe (Atkinson 2011; Bowerman 2011; Maddieson et al. 2011). The data thus encompass rich information on a large number of natural languages, allowing for the assignment of values to 127 immigrant languages in the NIS dataset. I use Atkinson's (2011) measure of phonemic diversity, which is constructed from raw WALS data on consonant phonemes, vowel phonemes, and tones. These phoneme inventories count the number of distinct, meaningful sounds in a language, and thus broadly represent the structural, sound-based complexity of a particular language. Following Atkinson (2011), I standardize the raw data on consonants, vowels, and tones and then average the values for each language to produce a measure of overall phonemic diversity.<sup>1</sup> I assign many of the languages missing WALS data the phonemic diversity values for the nearest language within the same language family.

Research in linguistics suggests that typological categories of language families are both conceptually and empirically important when analyzing the linguistic structures of the world's languages. The final models are estimated while controlling for language families. I construct

---

<sup>1</sup> My measure of phonemic diversity differs only slightly from Atkinson's in that I do not limit the sample to non-creole languages and instead include all languages during the standardization process.

individual dichotomous variables for the eight most attested language families in the sample. These language families encompass all of the most represented immigrant languages in the NIS dataset (see Table 2), as well as many other lesser-attested languages. All other languages are collapsed into an “Other” dichotomous variable. The reference group for these estimations comprises all Indo-European languages in the sample. All language family categorizations reflect the categories specified by WALS.

To test whether the accepted linguistic distance variable captures language background dynamics beyond language family, I include it in the full statistical models while controlling for language families. Values for this variable are taken from Chiswick and Miller (2005, pp. 5-6), and include all and only the extensions for additional languages that the authors include. The linguistic distance measure used by Chiswick and colleagues is the inverse of language scores provided by Hart-González and Lindemann’s (1993) technical report of foreign language training for the U.S. Department of State. Many attested languages in the NIS dataset do not have equivalent linguistic distance values. I assign such languages missing values for linguistic distance.

In addition to the main outcome variables and covariates, I control for variables that research suggests are important determinants of English proficiency among immigrants. In all statistical models, I control for female gender (female = 1), marriage to a U.S. citizen (married to citizen = 1), years of education, whether the respondent is employed or on temporary leave (employed or on leave = 1), age at migration, and years since migration. For the latter two controls, I include polynomial terms since effects have been found to be nonlinear (Akresh 2006; Chiswick and Miller 2008, 2010). In addition to years of education, there is reason to believe that distinct pre- and post-migration educational experiences might matter for English language proficiency. The NIS dataset provides a direct measure for each of these variables, and therefore I include the years of U.S. education for each respondent as well as whether the respondent took English classes before migrating to the U.S.<sup>2</sup>

In the literature on immigrant populations and language proficiency, Hispanic immigrants in the United States are often treated as a special group, typically with lower levels of overall proficiency in English. To control for these differences, I include a dichotomous variable representing respondents from Latin American countries.

The number of children in the household is also cited as an influence on adult immigrants’ English proficiency, although theories are often unclear as to the direction of this effect. Children could provide a disincentive to learning English as parents seek to foster native language knowledge within their own household. Alternatively, children (particularly school-aged children) could bring English into the household, thus increasing parents’ proficiency in English. I include a dichotomous variable for respondents with children less than 6 years of age (young children) and a dichotomous variable for respondents with children aged 6 to 17 (children).

Finally, the concentration of speakers of the same language within a specified area is found to negatively influence English proficiency. I control for language concentration by calculating a measure of the percentage of the respondent’s state population speaking the same language as the respondent. I do this for individuals who speak one of the 25 most common languages in the sample

---

<sup>2</sup> These variables range from being only slightly correlated with years of education (U.S. education,  $r = .19$ ) to being somewhat correlated with years of education (pre-migration English classes,  $r = .37$ ). Excluding these variables yields substantively identical results across all models.

while all other respondents are assigned a 0, assuming that language concentration in their states is negligible. Language data for individual states are from the U.S. Census 2000.

The relationships posited above are modeled using a logistic regression technique. Odds ratios are reported for all logit models, which represent the factor change in the odds that a respondent will speak English either well or very well. I then present predicted probabilities after reporting the odds ratios. Predicted probabilities provide a picture of changes in the probability of being proficient in English for certain sociodemographic groups while specifying specific values for substantively important variables such as phonemic diversity. To check the robustness of the main statistical models, I model the same relationships differently, first through binary probit models and then through ordered logistic regressions (not presented). Despite the various modeling techniques, all substantive findings from the main models I present remain identical.

The analysis proceeds in three steps. First, the effects of the linguistic variables—the *phonemic diversity* and *linguistic distance* of both the respondent’s childhood and home languages— and control covariates are estimated for both of the outcomes: speaking and understanding English (Table 3). This step estimates the relationship between the language background variables while controlling for individual, sociodemographic characteristics. Next, these same relationships are estimated while including language families (Table 4) to determine exactly which aspects of language background have a significant relationship with English proficiency among immigrants. Predicted probabilities are then presented for Hispanics and non-Hispanics by both gender and level of phonemic diversity (Table 5). The predictions represent the probabilities of speaking/understanding English for otherwise average speakers of Indo-European languages.

## Results/Findings

Table 1 presents descriptive statistics for the sample. One should note that respondents in the sample have overall higher levels of education compared to the broader immigrant population. This reflects the fact that the NIS sample is defined by legal permanent resident status, and therefore does not include undocumented immigrants or immigrants in other categories. The phonemic diversity scale, which is a standardized measure of the relative sound-based complexity of a language, varies from 0.87 to 1.13. The linguistic distance scale, which measures the “distance” between English and an immigrant’s native language via relative difficulty, ranges from 0.33 to 1. The language background measures have an overall correlation between  $r = 0.33 - 0.34$ .

**Table 1: Descriptive Statistics**

	Mean	Standard Deviation	Min.	Max.
<b>OUTCOMES:</b>				
Proficiency: Speaking English	0.50	0.50	0.00	1.00
Proficiency: Understanding English	0.57	0.50	0.00	1.00
<b>Explanatory Variables - Childhood Language</b>				
Phonemic Diversity	0.20	0.30	-0.87	1.13
Linguistic Distance	0.52	0.13	0.33	1.00
<b>Language families:</b>				
Indo-European	0.59	0.49	0.00	1.00



Afro-Asiatic	0.07	0.26	0.00	1.00
Austro-Asiatic	0.03	0.17	0.00	1.00
Austronesian	0.06	0.24	0.00	1.00
Sino-Tibetan	0.08	0.27	0.00	1.00
Dravidian	0.02	0.13	0.00	1.00
Korean	0.02	0.13	0.00	1.00
Japanese	0.01	0.09	0.00	1.00
Other	0.15	0.35	0.00	1.00
Explanatory Variables - Home Language				
Phonemic Diversity	0.21	0.30	-0.87	1.13
Linguistic Distance	0.52	0.13	0.33	1.00
Language families:				
Indo-European	0.69	0.46	0.00	1.00
Afro-Asiatic	0.06	0.24	0.00	1.00
Austro-Asiatic	0.03	0.16	0.00	1.00
Austronesian	0.05	0.21	0.00	1.00
Sino-Tibetan	0.07	0.26	0.00	1.00
Dravidian	0.01	0.11	0.00	1.00
Korean	0.02	0.13	0.00	1.00
Japanese	0.01	0.07	0.00	1.00
Other	0.09	0.29	0.00	1.00
Controls:				
Female	0.52	0.50	0.00	1.00
Education (Years)	12.55	4.93	0.00	21.00
U.S. Education (1=yes)	0.80	2.22	0.00	21.00
Employed (1=yes)	0.60	0.49	0.00	1.00
Married with American Spouse (1=yes)	0.08	0.28	0.00	1.00
Young Child (1=yes)	0.22	0.42	0.00	1.00
Child between 6-17 (1=yes)	0.30	0.46	0.00	1.00
Pre-Migration English Classes (1=yes)	0.38	0.49	0.00	1.00
Age at Migration	33.89	14.62	0.00	93.00
Years Since Migration	5.20	6.35	0.00	64.00
Language Concentration	6.67	10.18	0.00	28.75
Hispanic Ethnicity	0.35	0.48	0.00	1.00

Table 2 presents the most represented languages in the sample, along with their language families and phonemic diversity scores. Spanish is the most widely spoken childhood and home language among respondents. Other languages with over 200 speakers each include Tagalog, Mandarin Chinese, Russian, and Arabic. Although Indo-European languages are well represented in the sample, Austronesian, Sino-Tibetan, Afro-Asiatic, Austro-Asiatic, and Dravidian languages are also widely spoken.

**Table 2. Most Represented Language Families with Language Families and Phonemic Diversity Scores**

Language	Childhood Language (N)	Home Language (N)	Language Family	Phonemic Diversity
Spanish	2,454	2,261	Indo-European	0.13
Tagalog	341	257	Austronesian	-0.20
Mandarin Chinese	340	321	Sino-Tibetan	0.80
Russian	314	284	Indo-European	0.13
Arabic	257	225	Afro-Asiatic	-0.20
Vietnamese	189	181	Austro-Asiatic	0.80
Polish	182	172	Indo-European	0.13
Amharic	177	162	Afro-Asiatic	0.46
French	154	83	Indo-European	0.13
Cantonese	150	150	Sino-Tibetan	0.80
Gujarati	148	138	Indo-European	0.46
Korean	137	125	Korean	0.13
Haitian Creole a	129	107	Other	0.13
Hindi	126	96	Indo-European	0.46
Punjabi	111	94	Indo-European	0.46
Telugu	104	85	Dravidian	0.13
Albanian	78	72	Indo-European	0.13
Bengali	78	57	Indo-European	0.46
Urdu	77	70	Indo-European	0.46
Tamil	76	51	Dravidian	-0.20
Ukranian	75	65	Indo-European	0.13
Bulgarian	72	67	Indo-European	0.46
Malayalam	68	61	Dravidian	-0.20
Portuguese	67	51	Indo-European	0.13
Japanese	66	41	Japanese	-0.20
Romanian	63	51	Indo-European	0.13

Table 3 shows that language background works through both the language family and phonemic diversity. Notably, linguistic distance does not appear to significantly influence proficiency among immigrants in either speaking or understanding English. Models 2, 4, 6, and 8 show that higher levels of phonemic diversity lead to overall higher levels of proficiency in speaking and understanding English using both the childhood and home languages as explanatory variables. These results suggest that the phonemic complexity of a language may give speakers an advantage in learning English. As such, phonemic diversity can be considered a sort of linguistic resource that speakers draw on when learning to speak and understand English, net of an individual's sociodemographic characteristics. In terms of the childhood and home languages, results show that the home language is associated with overall stronger effects than the language spoken during

childhood. A one unit increase in the phonemic diversity of a current home language increases the odds of speaking English well or very well by a factor of 3.24. In contrast, a one-unit increase in the phonemic diversity of a childhood language increases the same odds by a factor of 3. A similar pattern is found with the likelihood of understanding English (an increase of factor 2.79 with the current home language versus an increase of factor 2.47 with the childhood language). These results suggest that the robust effects associated with phonemic diversity are felt more directly through the immigrant’s current home language than the childhood language.

Other estimates for the controls reported in Table 3 align with previous research. Years of education has a significant positive effect on both speaking and understanding English through both the childhood and home languages (increasing the odds of being proficient by a factor of at least 1.26). While arriving in the U.S. at an older age significantly decreases the odds of speaking English well or very well, staying in the U.S. longer significantly increases these odds. Both age at arrival and years since migration have curvilinear relationships with English proficiency, as their polynomial terms are also significant (not shown).<sup>3</sup> There is also a marked difference in the effect of gender on speaking and understanding English. Not only are females less likely to be proficient in English when compared to males, gender is found to negatively influence speaking English more than understanding English. The relationship between gender and proficiency leads to a decrease in the odds of proficiency in speaking English by a factor of at least 0.82 (versus a factor of at most 0.75 with understanding English). In terms of language families, it appears that speakers of an Austronesian language are much more likely to be proficient in English when compared to the reference group, where the odds of speaking English increases by a factor of 22.78 (based on home language). Other language families are associated with significant negative effects, including Austro-Asiatic languages, Sino-Tibetan languages, Japanese, and Korean. These results are similar to those found in previous studies when controlling for country of origin (see, e.g., Chiswick and Miller 1998a), though it is doubtful that countries of origin are fully comparable to language family groups.<sup>4</sup>

**Table 3: Estimates of English Proficiency Among Adult Immigrants by Non-English Language Background: Lang. Spoken During Childhood and Lang. Currently Spoken at Home (Odds Ratios)**

	SPEAK ENGLISH				UNDERSTAND ENGLISH			
	Childhood Language		Home Language		Childhood Language		Home Language	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Phonemic Diversity		3.00*** (6.74)		3.24*** (6.42)		2.47*** (5.49)		2.79*** (5.61)
Linguistic Distance	2.28 (1.26)		2.66 (1.33)		1.52 (0.61)		1.58 (0.61)	
<b>LANG. FAMILIES: <sup>a</sup></b>								
Afro-Asiatic	1.12 (0.75)	1.68*** (3.98)	1.21 (1.20)	1.90*** (4.64)	1.05 (0.29)	1.43** (2.68)	1.16 (0.92)	1.63*** (3.48)

<sup>3</sup> Supplementary analyses for males with average education show that after 15 years old, English proficiency decreases at an increasing rate with age at migration until around 70 years old, when the negative influence levels off. The same analyses show that years since migration exerts an increasingly positive influence on English proficiency until around 20 to 30 years since migration, when the rate of such influence levels off.

<sup>4</sup> Regressing reported countries of birth on the full range of language family variables reveals that only 13-15% of the variance in country of birth can be explained by language family groups.

Austro-Asiatic	0.14*** (-6.49)	0.11*** (-7.68)	0.18*** (-5.35)	0.14*** (-6.46)	0.09*** (-7.83)	0.07*** (-9.00)	0.10*** (-6.82)	0.08*** (-8.03)
Austronesian	9.96*** (13.46)	17.09*** (15.58)	13.10*** (13.52)	22.78*** (15.43)	12.16*** (12.96)	19.15*** (14.57)	15.31*** (12.92)	23.39*** (14.37)
Sino-Tibetan	0.33*** (-5.57)	0.23*** (-8.48)	0.36*** (-4.54)	0.24*** (-7.34)	0.34*** (-5.34)	0.24*** (-8.22)	0.38*** (-4.32)	0.25*** (-7.40)
Korean	0.11*** (-5.16)	0.20*** (-6.27)	0.12*** (-4.53)	0.24*** (-5.37)	0.25** (-3.28)	0.36*** (-4.24)	0.29** (-2.70)	0.42*** (-3.42)
Japanese	0.19*** (-3.58)	0.48* (-2.25)	0.19** (-2.98)	0.54 (-1.49)	0.30* (-2.48)	0.57 (-1.62)	0.30* (-2.14)	0.62 (-1.16)
Other	2.33*** (5.14)	2.32*** (6.85)	2.62*** (5.04)	2.10*** (5.19)	2.02*** (4.07)	1.93*** (5.15)	2.12*** (3.78)	1.55** (3.03)
Dravidian <sup>b</sup>	1.66 (0.48)	3.77** (2.72)	-- --	5.90** (3.27)	-- --	7.03** (2.98)	-- --	9.84*** (3.40)
<b><i>Sociodemographics:</i></b>								
Female	0.83* (-2.43)	0.83** (-2.66)	0.83* (-2.28)	0.82* (-2.52)	0.73*** (-4.10)	0.73*** (-4.31)	0.74*** (-3.81)	0.75*** (-3.84)
Education	1.26*** (20.34)	1.28*** (22.31)	1.26*** (18.67)	1.28*** (20.41)	1.26*** (20.58)	1.28*** (22.48)	1.26*** (19.27)	1.27*** (20.90)
U.S. Education	1.49*** (11.02)	1.47*** (10.98)	1.47*** (9.73)	1.46*** (9.88)	1.49*** (9.20)	1.46*** (9.09)	1.43*** (7.94)	1.42*** (8.07)
Employed	1.43*** (4.36)	1.35*** (3.85)	1.35*** (3.39)	1.33*** (3.33)	1.28** (2.99)	1.21* (2.43)	1.29** (2.91)	1.29** (3.06)
Married, American Spouse	2.45*** (6.10)	2.64*** (6.74)	1.31 (1.39)	1.39 (1.73)	2.68*** (6.40)	2.88*** (6.99)	1.60* (2.48)	1.67** (2.75)
Young Child	1.10 (1.09)	1.14 (1.53)	1.15 (1.48)	1.22* (2.23)	1.21* (2.11)	1.22* (2.37)	1.23* (2.18)	1.30** (2.93)
Child	0.90 (-1.36)	0.84* (-2.17)	0.98 (-0.21)	0.92 (-1.04)	0.88 (-1.55)	0.84* (-2.25)	0.94 (-0.78)	0.89 (-1.50)
Pre-Migration English Classes	2.05*** (9.47)	2.07*** (10.10)	2.12*** (9.24)	2.09*** (9.45)	2.00*** (8.86)	2.03*** (9.55)	2.04*** (8.66)	2.02*** (8.95)
Age at Migration <sup>c</sup>	0.92*** (-5.16)	0.91*** (-6.08)	0.91*** (-5.89)	0.91*** (-6.16)	0.93*** (-4.69)	0.92*** (-5.78)	0.92*** (-5.26)	0.92*** (-5.65)
Years since Migration <sup>c</sup>	1.20***	1.20***	1.23***	1.23***	1.21***	1.21***	1.23***	1.23***

	(11.97)	(12.43)	(10.47)	(11.15)	(12.85)	(13.14)	(11.24)	(11.86)
Language Concentration	0.96***	0.96***	0.96***	0.96***	0.97***	0.97***	0.97***	0.97***
	(-6.39)	(-6.22)	(-4.71)	(-4.66)	(-4.80)	(-4.75)	(-3.64)	(-3.63)
N	6003	6607	5066	5521	5984	6607	5066	5521
pseudo R <sup>2</sup>	0.403	0.406	0.381	0.386	0.399	0.403	0.375	0.377

<sup>a</sup> Reference category: Indo-European language family

<sup>b</sup> Dravidian language family not included in indicated models due to perfect predictions when estimated with linguistic distance variable.

<sup>c</sup> Polynomial terms included in all estimations (not shown).

Odds ratios reported; z statistics in parentheses

Source: New Immigrant Survey 2003, Restricted Dataset v. 1; language data collected from World Atlas of Language Structures

Odds ratios reported; z statistics in parentheses

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

To show how these relationships influence proficiency for key groups, Table 4 presents predicted probabilities for prototypical individuals speaking Indo-European languages. I set values for all predictions to reflect the following sociodemographic characteristics: average levels of education, average levels of U.S. education, no English classes before coming to the U.S., currently employed or on temporary leave, average age at arrival to the U.S., the average number of years since migration, no children, and residence in an area with average levels of linguistic concentration. I then manipulate values for gender, Hispanic versus non-Hispanic ethnicity, and levels of phonemic diversity for both childhood and home languages.<sup>5</sup>

Results from Table 4 thus show the predicted probabilities of speaking or understanding English well/very well along these three dimensions. Going from low levels of phonemic diversity to high levels of phonemic diversity consistently increases one's likelihood of speaking and understanding English. The highest gains associated with increased phonemic diversity are found with the influence of language background on *speaking* English. For all subgroups, going from a *childhood language* with low phonemic diversity to one with high phonemic diversity yields an increase in the probability of speaking English ranging from .33 (among female Hispanics) to .35 (among female non-Hispanics). Changes in the phonemic diversity of the *home language* produce similar results: increases from .30 (among female Hispanics) to .37 (among male non-Hispanics). The lowest increases are found with the effect of childhood language on *understanding* English. Going from a childhood language with low phonemic diversity to one with high phonemic diversity produces an increase in the probability of understanding English ranging from .26 (among male non-Hispanics) to .29 (among female and male Hispanics). Thus, it appears that increases in phonemic diversity yield the highest gains with regard to speaking proficiency, and the lowest gains in understanding English. In all realms, the phonemic diversity advantage benefits males more than females, producing higher probabilities of both speaking and understanding English for males.

<sup>5</sup> Hispanic ethnicity is presented in the interest of comprehensiveness since it is consistently found to be a key sociodemographic cleavage with respect to English proficiency. Practically speaking, however, most Hispanics are likely to speak Spanish, which is a language with an average level of phonemic diversity. There are, however, a number of Hispanics who speak indigenous languages, which could differ in phonemic diversity.

**Table 4: Predicted Probabilities of English Proficiency by Levels of Phonemic Diversity Among Average Immigrant Speakers of an Indo-European Language**

	Speak English		Understand English	
	Childhood Language	Home Language	Childhood Language	Home Language
<b>Female Hispanic</b>				
Low Phonemic Diversity	0.24	0.15	0.33	0.24
Average Phonemic Diversity	0.31	0.20	0.40	0.30
High Phonemic Diversity	0.57	0.45	0.62	0.55
<b>Female Non-Hispanic</b>				
Low Phonemic Diversity	0.35	0.24	0.48	0.36
Average Phonemic Diversity	0.44	0.32	0.55	0.44
High Phonemic Diversity	0.70	0.60	0.75	0.69
<b>Male Hispanic</b>				
Low Phonemic Diversity	0.27	0.17	0.40	0.29
Average Phonemic Diversity	0.34	0.23	0.48	0.37
High Phonemic Diversity	0.61	0.50	0.69	0.62
<b>Male Non-Hispanic</b>				
Low Phonemic Diversity	0.40	0.28	0.55	0.43
Average Phonemic Diversity	0.49	0.36	0.63	0.52
High Phonemic Diversity	0.74	0.65	0.81	0.75

I conduct supplemental analyses to assess the robustness of the findings across samples and to determine whether language background interacts with any of the sociodemographic variables (results not shown). Because the samples differ somewhat according to which language variable is estimated, I estimate additional models for both home language and childhood language that

include all language variables: linguistic distance, phonemic diversity, and language families. This allows me to test language-based relationships using the same sample. In all additional models, linguistic distance fails to show any significant relationship with English proficiency. In contrast, results show that the phonemic diversity of both the home language and childhood language significantly influences speaking and understanding English, with identical patterns as those presented in Table 4. These supplementary analyses increase confidence in the findings since they do not appear to be artifacts of samples differences.

Years of education appear to interact with phonemic diversity in all models for home language and in the model estimating the relationship between childhood language and understanding English (odds ratio for the latter model = 1.10,  $z = 2.75$ ,  $p < .01$ ). This suggests that the influence of phonemic diversity is dependent on the years of education an immigrant has. Further analysis calculating marginal effects reveals that the significant influence of language background is felt after approximately 9 years of education, with the magnitude of such influence peaking at 16 years of education. Only one other interaction with language background is statistically significant at conventional levels: the influence of home language on speaking English does not surface for those married to an American spouse (odds ratio for interaction term = 0.07,  $z = -2.10$ ,  $p < .05$ ). This suggests that English may be dominant in such bilingual households, making a language background effect trivial. Aside from education and marriage to an American spouse, it appears that language background works largely independent of other sociodemographic dynamics.

## Limitations/challenges

Despite the robust findings presented, there are, of course, a number of limitations with the current study. As the sample includes new legal permanent residents only, it necessarily excludes more broadly defined immigrant populations. Undocumented immigrants, for example, are not included in the estimations presented above. Although there is no *a priori* reason to believe that linguistic structures would affect proficiency among these excluded populations any differently, this is a hypothesis that must be tested with other datasets.

Furthermore, language as an object of analysis is complex and, as such, this study only purports to uncover one aspect of immigrants' native language characteristics. The results presented above focus on the *phonological* aspects of language differences between immigrant groups. However, there are a number of facets of language that are necessarily excluded from the analysis. For instance, although phonemic diversity appears to increase the probability of English proficiency, differences in the *syntactic* structures of English and immigrant languages may lead to a decrease in proficiency. Syntactic structures refer to the *ordering* of different components of language, such as the subject, object, and verb of different phrases. Future research on immigrants' language backgrounds should therefore focus on the multiple aspects of language and pinpoint linguistic dynamics net of other structural effects.

## Policy recommendations

The findings from this study lend themselves easily to policy recommendations. At the heart of the research is whether immigrant groups are positioned differently according to their linguistic backgrounds in their acquisition of key English language skills and, consequently, in their integration into American society. For immigrants who do not speak English, acquiring adequate levels of

linguistic skills imposes costs, such as the resources to pay for language instruction and materials, and a significant commitment of time. Offsetting these costs through policies that facilitate language learning among non-native populations could lead to higher mean levels of English language proficiency and, ultimately, more effective integration. For example, providing grants to qualified English as a second language schools or educational institutions could help offset the costs that immigrants learning English would otherwise have to bear themselves. Other developed countries have a history of formulating policies that facilitate language training for immigrant populations (Norway, Sweden, Denmark, Switzerland, for example), and have had some success with higher rates of language acquisition. It would also be possible to pinpoint which language groups need additional support. Stated in terms of the study's findings, if certain immigrant groups are at a comparative disadvantage when learning English due to more or less phonemic diversity, steps could be taken to address this disparity by identifying the immigrant groups that need the most assistance due to their linguistic backgrounds. Resources could then be appropriated that target these groups first or primarily when learning English.

Such findings could also guide policy regarding state-immigrant contact. One way this line of research could inform bureaucratic policy and practices is by facilitating communication in interactions such as immigrant interviews. If certain immigrant groups have overall lower levels of English proficiency upon arrival in the U.S. as influenced by their linguistic backgrounds, interpreter services might be required in their interactions with USCIS personnel, for example. The interpreters employed by USCIS could also benefit from knowledge about the phonemic diversity of the immigrant's language. According to findings, immigrants speaking languages that are less phonemically complex should acquire proficiency in English at a lower or slower rate, all else equal. By targeting the immigrant groups that are more likely to need language assistance for interview purposes, limited resources dedicated to interpretation or translation of documents can be allocated more efficiently.

Another area of state-immigrant contact that might be improved based on the findings from this research includes customer service. The USCIS has as one of its missions to provide effective customer-oriented immigration benefit and information services. In order to accomplish this mission, adequate resources and materials that are useful to new immigrants are necessary. Of course, making such materials available in non-English languages for those new immigrants with limited English proficiency ensures these materials' usefulness. Findings from this research could help in deciding which non-English languages should be a priority when producing and providing these materials and resources. If, for example, Arabic and Tamil have low levels of phonemic diversity leading to overall lower levels of English proficiency on average, it would be appropriate to provide more customer-based materials and resources in these languages. Clearly, such determinations would also vary by the size of the immigrant populations served. Thus, the findings outlined above could aid in allotting limited resources for the goal of effective customer service.

## Next steps in research

There are multiple avenues for future research on immigrants' language proficiency. As the findings outlined above are limited to the *phonological* differences in language, future studies might focus on other aspects of language structure. For example, there may be important structural differences between immigrant languages and English along the lines of linguistic syntax, semantics, or morphology that offset or increase the effects that this study finds. It may be possible, for example, that syntactic complexity—or the complexity of the rules by which phrases in a language are constructed—influences levels of English proficiency as well. New lines of research can be forged



that take into account these linguistic differences and attempt to isolate structural effects net of the sociodemographic differences between immigrant populations.

With future research and analysis, it may also be possible to isolate the effects of the first generation's English proficiency on important second-generation outcomes, either linguistic or otherwise. The examination of this intergenerational transfer of linguistic capital could have important implications for how integration policies are formulated for the second generation and beyond. Effects that go beyond the first generation cohort are particularly meaningful within the fields of bilingual education and stratification. For example, bilingual language programs may be less successful when they assume that all immigrant language groups are similarly positioned. Certain linguistic groups may need additional time to achieve the levels of proficiency required to be categorized as "proficient bilinguals", i.e., proficient in both English and the non-English language. Not recognizing these differences puts educators in the position of assessing proficiency without being aware of language background differences. Such assessments could lead to stratification among second-generation immigrant children, better known as "tracking" in education. In sum, any residual effects of the first generation's linguistic differences on their offspring could better inform policy makers on how to trace development and progress across time based on fundamental linguistic differences.

With the future available of NIS data, research should begin to focus on timed trends associated with language variables such as proficiency and language background. The findings outlined above apply to new immigrants and their levels of language proficiency at one point in time. However, with new panel data, researchers can begin to identify trends in language proficiency over time. Possible research questions include the following: Do immigrants who speak languages with higher phonemic diversity learn English at a faster rate? How is this affected by other important time-specific variables, such as years since migration and age at migration? This line of research could provide fruitful insight into the dynamic processes of English language proficiency among immigrants.

Future research should also focus on the effect of language background on other important integration outcomes. It is quite possible that language background has an effect on the economic or political integration of new immigrants in America. Now that language background has been identified as a significant predictor of linguistic integration, future studies should test whether this is the case for economic integration outcomes, such as labor market integration, or political outcomes, such as voting and other forms of civic participation. Future analyses should take care to isolate such potential language background effects from the influence associated with self-reported language proficiency. Uncovering additional language background effects could give us a better understanding of the ways non-sociodemographic variables influence immigrant integration in both the short and long terms.

## **Works cited**

Akresh, I.R., 2006. Contexts of English language use among immigrants to the United States. *International Migration Review* 41, 930–955.

Akresh, I.R., 2007a. Dietary assimilation and health among Hispanic immigrants to the United States. *Journal of Health and Social Behavior* 48, 404–417.

Akresh, I.R., 2007b. U.S. immigrants' labor market adjustment: Additional human capital investment and earnings growth. *Demography* 44, 865–881.

- Akresh, I.R., 2008. Occupational trajectories of legal US immigrants: Downgrading and recovery. *Population and Development Review* 34, 435–456.
- Alba, R., Nee, V., 2003. *Remaking the American Mainstream: Assimilation and Contemporary Immigration*. Harvard University Press, Cambridge, MA.
- Arcia, E., Skinner, M., Bailey, D., Correa, V., 2001. Models of acculturation and health behaviors among Latino immigrants to the US. *Social Science & Medicine* 53, 41–53.
- Atkinson, Q.D., 2011. Phonemic diversity supports a serial founder effect model of language expansion from Africa. *Science* 332, 346–349.
- Bakker, D., Müller, A., Velupillai, V., Wichmann, S., Brown, C.H., Brown, P., Egorov, D., Mailhammer, R., Grant, A., Holman, E.W., 2009. Adding typology to lexicostatistics: A combined approach to language classification. *Linguistic Typology* 13, 169–181.
- Bean, F.D., Stevens, G., 2003. *America's Newcomers and the Dynamics of Diversity*. Russell Sage Foundation, New York, NY.
- Beenstock, M., Chiswick, B.R., Repetto, G.L., 2001. The effect of linguistic distance and country of origin on immigrant language skills: Application to Israel. *International Migration* 39, 33–60.
- Bleakley, H., Chin, A., 2004. Language skills and earnings: Evidence from childhood immigrants. *The Review of Economics and Statistics* 86, 481–496.
- Bleakley, H., Chin, A., 2008. What holds back the second generation?: The intergenerational transmission of language human capital among immigrants. *Journal of Human Resources* 43, 267–298.
- Bloemraad, I., 2002. The North American naturalization gap: An institutional approach to citizenship acquisition in the United States and Canada. *International Migration Review* 36, 193–228.
- Bloemraad, I., 2006. *Becoming a Citizen: Incorporating Immigrants and Refugees in the United States and Canada*. University of California Press, Berkeley, CA.
- Bowern, C., 2011. Out of Africa? The logic of phoneme inventories and founder effects. *Linguistic Typology* 15, 207–216.
- Bradlow, A.R., Bent, T., 2008. Perceptual adaptation to non-native speech. *Cognition* 106, 707–729.
- Casey, T., Dustmann, C., 2008. Intergenerational transmission of language capital and economic outcomes. *Journal of Human Resources* 43, 660–687.
- Chiswick, B.R., 1991. Speaking, reading, and earnings among low-skilled immigrants. *Journal of Labor Economics* 9, 149–170.
- Chiswick, B.R., 1998. Hebrew language usage: Determinants and effects on earnings among immigrants in Israel. *Journal of Population Economics* 11, 253–271.

- Chiswick, B.R., Miller, P.W., 1995. The endogeneity between language and earnings: International analyses. *Journal of Labor Economics* 13, 246–288.
- Chiswick, B.R., Miller, P.W., 1998a. English language fluency among immigrants in the United States. *Research in Labor Economics* 17, 151–200.
- Chiswick, B.R., Miller, P.W., 1998b. Language skill definition: A study of legalized aliens. *International Migration Review* 32, 877–900.
- Chiswick, B.R., Miller, P.W., 2001. A model of destination-language acquisition: Application to male immigrants in Canada. *Demography* 38, 391–409.
- Chiswick, B.R., Miller, P.W., 2005. Linguistic distance: A quantitative measure of the distance between English and other languages. *Journal of Multilingual & Multicultural Development* 26, 1–11.
- Chiswick, B.R., Miller, P.W., 2007a. *The Economics of Language: International Analyses*. Routledge, New York, NY.
- Chiswick, B.R., Miller, P.W., 2007b. Linguistic distance: A quantitative measure of the distance between English and other languages, in: [252007aChiswick and Miller]. chapter 20. pp. 575–584.
- Chiswick, B.R., Miller, P.W., 2008. Modeling immigrants' language skills. *Research in Labor Economics* 27, 75–128.
- Chiswick, B.R., Miller, P.W., 2010. Occupational language requirements and the value of English in the U.S. labor market. *Journal of Population Economics* 23, 353–372.
- Cho, W.K.T., 1999. Naturalization, socialization, participation: Immigrants and (non-)voting. *The Journal of Politics* 61, 1140–1155.
- Dávila, A., Mora, M.T., 2000. English skills, earnings, and the occupational sorting of Mexican Americans along the U.S.-Mexico border. *International Migration Review* 34, 133–157.
- Dryer, M., Haspelmath, M. (Eds.), 2011. *The World Atlas of Language Structures Online*. Max Planck Digital Library, Munich.
- Dulay, H., Burt, M., Krashen, S., 1982. *Language Two*. Oxford University Press, New York, NY.
- Dustmann, C., 1994. Speaking fluency, writing fluency and earnings of migrants. *Journal of Population Economics* 7, 133–156.
- Dustmann, C., van Soest, A., 2002. Language and the earnings of immigrants. *Industrial and Labor Relations Review* 55, 473–492.
- Espenshade, T.J., Fu, H., 1997. An analysis of English-language proficiency among U.S. immigrants. *American Sociological Review* 62, 288–305.
- Espinosa, K.E., Massey, D.S., 1997. Determinants of English proficiency among Mexican migrants to

the United States. *International Migration Review* 31, 28–50.

- Gathercole, V.C.M., 2007. Miami and North Wales, so far and yet so near: A constructivist account of morphosyntactic development in bilingual children. *International Journal of Bilingual Education and Bilingualism* 10, 224–247.
- Gilbert, G., 1983. Transfer in second language acquisition, in: Andersen, R.W. (Ed.), *Pidginization and Creolization as Language Transfer*. Newbury House, Rowley, MA, pp. 168–181.
- Grenier, G., 1984a. The effects of language characteristics on the wages of Hispanic-American males. *The Journal of Human Resources* 19, 35–52.
- Grenier, G., 1984b. Shifts to English as usual language by Americans of Spanish mother tongue. *Social Science Quarterly* 65, 537–550.
- Hart-González, L., Lindemann, S., 1993. *Expected Achievement in Speaking Proficiency, 1993*. Technical Report. School of Language Studies, Foreign Services Institute, U.S. Department of State.
- Ioup, G., Weinberger, S.H., 1987. *Interlanguage Phonology: The Acquisition of a Second Language Sound System*. Newbury House, Cambridge, MA.
- Jasso, G., 2011. Migration and stratification. *Social Science Research* 40, 1292–1336.
- Jasso, G., Rosenzweig, M.R., 1990. *The New Chosen People: Immigrants in the United States*. Russell Sage Foundation, New York.
- Kossoudji, S.A., 1988. English language ability and the labor market opportunities of Hispanic and East Asian immigrant men. *Journal of Labor Economics* 6, 205–228.
- Lev-Ari, S., Keysar, B., 2010. Why don't we believe non-native speakers? the influence of accent on credibility. *Journal of Experimental Social Psychology* 46, 1093–1096.
- Lindstrom, D.P., Massey, D.S., 1994. Selective emigration, cohort quality, and models of immigrant assimilation. *Social Science Research* 23, 315–349.
- Maddieson, I., Bhattacharya, T., Smith, D.E., Croft, W., 2011. Geographical distribution of phonological complexity. *Linguistic Typology* 15, 267–279.
- McManus, W., Gould, W., Welch, F., 1983. Earnings of Hispanic men: The role of English language proficiency. *Journal of Labor Economics* 1, 101–130.
- McManus, W.S., 1985. Labor market costs of language disparity: An interpretation of Hispanic earnings differences. *The American Economic Review* 75, 818–827.
- Odling, T., 1989. *Language Transfer*. Cambridge University Press, Cambridge, U.K.
- Paradis, J., 2010. Bilingual children's acquisition of English verb morphology: Effects of language exposure, structure complexity, and task type. *Language Learning* 60, 651–680.

- Portes, A., Rumbaut, R.G., 2006. *Immigrant America: A Portrait*. University of California Press, Berkeley, CA. 3rd edition.
- Purcell, E.T., Suter, R.W., 1980. Predictors of pronunciation accuracy: A reexamination. *Language Learning* 30, 271–287.
- Romaine, S., 2003. Variation, in: Doughty, C.J., Long, M.H. (Eds.), *The Handbook of Second Language Acquisition*. Blackwell, Malden, MA. chapter 14, pp. 409–435.
- Shields, M.A., Price, S.W., 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.
- Stevens, G., 1992. The social and demographic context of language use in the United States. *American Sociological Review* 57, 171–185.
- Stevens, G., 1999a. Age at immigration and second language proficiency among foreign-born adults. *Language in Society* 28, 555–578.
- Stevens, G., 1999b. A century of U.S. censuses and the language characteristics of immigrants. *Demography* 36, 387–397.
- Suter, R.W., 1976. Predictors of pronunciation accuracy in second language learning. *Language Learning* 26, 233–253.
- Tainer, E., 1988. English language proficiency and the determination of earnings among foreign-born men. *The Journal of Human Resources* 23, 108–122.
- Wichmann, S., Müller, A., Velupillai, V., 2010. Homelands of the world's language families: A quantitative approach. *Diachronica* 27, 247–276.
- Xi, J., Hwang, S.S., Cao, Y., 2010. Ecological context and immigrants' earnings: English ability as a mediator. *Social Science Research* 39, 652–661.