ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Chiswick, Barry R.; Gindelsky, Marina

Working Paper

Determinants of Bilingualism among Children

IZA Discussion Papers, No. 8488

Provided in Cooperation with:
IZA Network @ LISER, Luxembourg Institute of Socio-Economic Research (LISER)

Suggested Citation: Chiswick, Barry R.; Gindelsky, Marina (2014) : Determinants of Bilingualism among Children, IZA Discussion Papers, No. 8488, Institute for the Study of Labor (IZA), Bonn

This Version is available at:
https://hdl.handle.net/10419/103540

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

DISCUSSION PAPER SERIES

IZA DP No. 8488

Determinants of Bilingualism among Children

Barry R. Chiswick
Marina Gindelsky

September 2014

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Determinants of Bilingualism among Children

Barry R. Chiswick
George Washington University
and IZA

Marina Gindelsky
George Washington University

Discussion Paper No. 8488
September 2014

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

IZA Discussion Paper No. 8488
September 2014

ABSTRACT

Determinants of Bilingualism among Children*

This paper analyzes the determinants of bilingualism (i.e., speaks a language other than English at home) among children age 5 to 18 years in the American Community Survey, 2005-2011. Two groups of children are considered: those born in the US (native born) and foreign-born children who immigrated prior to age 14 (the 1.5 generation). The analyses are conducted overall, within genders, and within racial and ethnic groups. Bilingualism is more prevalent if the parents are foreign born, less proficient in English, of the same ancestry (linguistic) group, and if the child lives in an ethnic (linguistic) concentration area. Although the effects are relatively smaller, a foreign-born grandparent living in the household increases child bilingualism, while a higher level of parental education tends to decrease it. Children of Asian and especially of Hispanic origin are more likely to be bilingual than their white, non-Hispanic counterparts. Native-born Indigenous children are more likely to be bilingual.

JEL Classification: J15, J24, I210, Z13

Keywords: bilingualism, native born children, immigrant children, family

Corresponding author:

Barry R. Chiswick
Department of Economics
George Washington University
2115 G Street, NW
Monroe Hall 340
Washington, DC 20052
USA
E-mail: brchis@uic.edu

* Paper presented at "Economy and Language: An Inter-Disciplinary Workshop", University of Chicago Center in Paris, June 2014.

Determinants of Bilingualism Among Children
Barry R. Chiswick and Marina Gindelsky

I. Introduction

Little attention has been devoted to the determinants of bilingualism among children in the United States. This is largely due to the lack of suitable data. However, the origins of bilingualism among native-born children is a topic of particular interest. Usually, considerable investment in learning a second language (a form of human capital) is required as an adult; however, children do not usually make such an investment consciously and may obtain their knowledge through another path, such as living with a parent, or other caregiver, who speaks another language at home.

It is also important to consider foreign-born children who immigrated at a very young age, for example, prior to the age of 14, or, the "1.5 generation". These children may have a natural capacity for being bilingual due to pre-migration exposure to the language in the country of origin, but they do not necessarily practice their bilingualism at home.

Immigrant or linguistic minority parents (or other relatives) may want to maintain their cultural heritage by speaking the origin language at home, or conversely, encourage assimilation by speaking only the dominant language (English in the case of the United States) in the home. Additionally, bilingual practices among children are likely to vary by gender, ethnicity, and household structure, among other determinants.

This paper uses the American Community Survey, Public Use Microdata Sample (2005-2011), as reconstructed by IPUMS from the Minnesota Population Center, to construct a child-

level dataset and explore the determinants of bilingualism for children age 18 years and younger who are living at home, and who were either born in the U.S. or immigrated before the age of 14. Section II reviews the existing literature on advantages of bilingualism and potential factors which may influence childhood bilingualism. Section III discusses the data and develops a model. Section IV presents and discusses the results of a multivariate statistical analysis of the determinants of bilingualism among children in the U.S. and Section V summarizes and concludes.

II. Literature Review

In order to understand the determinants of bilingualism among children, we first consider whether there is an economic advantage to being a bilingual adult. There have been several studies on this topic which analyze the advantages (or disadvantages) of being bilingual in the labor market.

Knowledge of a second language, in addition to the dominant language, may not only increase potential earnings but also expand possible job market opportunities. An earnings premium may be due to ever globalizing markets which generate demand among employers for US workers with fluency in languages other than English (Saiz and Zoido 2002). Additionally, hiring workers with knowledge of a locally prevalent foreign language (for example, Spanish in a Hispanic enclave) may provide a comparative advantage for firms offering goods and services within the ethnic community. For both men and women, knowledge of a second language may function as a signal of "advanced cognitive and communicative abilities" (Saiz and Zoido 2002).

Moreover, fluency in multiple languages can have an important impact on cognitive development. In a study examining the effects of bilingualism on cognitive aging Bak et al.

(2014) found a "protective effect" of bilingualism against cognitive decline in the elderly independently of childhood intelligence, controlling for immigrants status, gender, and soceioeconomic status. However, individuals with higher levels of intelligence may benefit more from early acquisition, while those of lower intelligence from later acquisition. Most interestingly, Bak et al. (2014) found that knowing 3 or more languages was more beneficial than knowing two, while the degree to which the second language was used in daily life made little difference in the cognitive returns.

However, if knowledge of a second language detracts from proficiency in English, particularly in a job in which a high level of English skill is essential, it could result in an earnings penalty (Chiswick and Miller 2007). The presence of a noticeable "accent" while speaking or writing English may result in an earnings penalty either directly due to difficulty in communication or as a signal of an immigrant or ethnic minority ancestry, particularly if discrimination is present (Chiswick and Miller 2013).

One of the most important factors in determining the advantages of bilingualism relates to the demand for the second language in the local environment. This demand may arise from an earnings premium in the local labor market or from opportunities to participate in the community as a whole. Using Canadian 1991 Census data for the metropolitan areas of Montreal, Toronto, and Vancouver, Pendakur and Pendakur (2002) find that, in general, knowledge of a minority language is associated with *lower* earnings, underscoring the importance of market demand. While being bilingual in the two dominant languages in Canada (French/English) helps in Montreal, it hurts in other areas. Regarding other languages, the differentials are not large for women, but men in Canada who speak Spanish, Arabic, and Tagalog earn less than those who know only English in the three major cities (Pendakur and Pendakur 2002). Although their focus

is not on the comparison with native-born Canadians, the results emphasize the importance of evaluating differences by ethnic group and gender.

The question remains: If bilingualism has an ambiguous effect on labor market earnings, is there an economic incentive for children to learn to speak another language at home? Perhaps speaking English without an accent that might reveal another language proficiency allows them as adults to capture any economic benefits that exist, while avoiding discrimination against foreign language speakers (Chiswick and Miller 2013).

In addition to considering the mixed results on labor market returns, we can also consider additional benefits which may arise from bilingualism. There have been inquiries regarding the effects of both an immigrant background and bilingualism on educational outcomes. In her paper, Feliciano (2001) shows that immigrants are less likely to drop out of school and demonstrates that bilingual youths are more likely to be bicultural. She documents that there is a strong cultural component in maintaining bilingualism which may play a role in educational success. Portes and Rivas (2011) review the effects of differences in human capital stemming from migrant skill or legal status on linguistic and cultural assimilation, education, and earnings outcomes, and conclude that preserving linguistic and cultural heritage is actually advantageous for immigrant children. But how does this language preservation happen?

In her discussion of language acquisition, Paradis describes the development of simultaneous and sequential bilingualism in children (Paradis 2007). Children may become simultaneous bilinguals if they acquire both languages in the home, or in the daycare, or they may become sequentially bilingual wherein they speak their first language at home and the

second at school. Generally, children who have immigrated become sequentially bilingual and often one language is more dominant than another (Paradis 2007).

Which language emerges as dominant may have important effects on future labor market and educational outcomes. The cognitive abilities required for maintaining two languages may be advantageous for older children in both education and in the labor market, especially due to practice with conceptual-lexical mappings. Paradis also states, however, that the vocabulary in each language in typically smaller for bilingual children than for monolingual children. If their English vocabulary suffers as a result, the second language may have been a hindrance, particularly when the two languages have substantially different grammatical structures. A certain amount of mixing of languages inevitably occurs and children must choose which language to speak whether in the home, at school, or among their peers. Studying bilingualism in children is of particular importance because by age 3, bilingual children "display rule-governed and adult-like patterns" (Paradis 2007). Moreover, the extent to which children retain their bilingualism as adults may depend not only on the environment in the home, but also on opportunities for using the immigrant language in the community, especially in an ethnic enclave.

Biedinger et al. (2014) attempt to answer the question of which contexts of exposure matter most in the retention of a heritage language for children, using data on young migrant Turkish children in Germany. They find that all contexts are relevant at different stages of development for retention of the language, but also that the parental input/family context is the most significant factor, both in parental Turkish-language ability (positive effect) and in the education of the care giver (negative effect). Moreover, both exposure to German language

media and attending a German-speaking preschool have negative effects on the maintenance and proficiency of Turkish.

Feng et al. (2014) analyze the determinants of early reading among bilingual children and conclude that they are at a disadvantage in that they tend to live in households with fewer resources and fewer family activities that promote early reading. Chiswick, Lee and Miller (2005) examine learning in the home and find that among immigrant households English language proficiency among children is enhanced the greater the English proficiency of their parents, with the mother's proficiency more important than that of the father.

Although the literature is inclusive regarding the effects of bilingualism on educational and earnings outcomes, this paper attempts to understand why and how children become bilingual. By constructing a child-level dataset from survey data, this paper identifies several important factors determining whether a child speaks a foreign language at home. It will also examine the effects on child bilingualism of gender, ethnicity, and nativity.

III. Data and Model

The analyses in this paper are based on the 2005-2011 American Community Survey (ACS). This is a monthly survey comprising all 50 states and the District of Columbia and is released in the form of a Public Use Microdata Sample (reconstructed by IPUMS). This survey was chosen for this analysis due to the wide range of demographic variables available, in addition to the large sample size, which facilitates analyses for very specific language groups.

Language questions are asked in the ACS only for persons age 5 and over. As a result, the sample under study is for boys and girls living at home ages 5 to 18[1]. The sample is further restricted to children who are either native born, or immigrated to the United States prior to the age of 14. One of the more difficult tasks of this study is identifying children in a household structure and attributing to them other family or household characteristics. Respondents to survey questions in the ACS must be someone living in the household over the age of 15, and this person can respond on behalf of a younger child or other relative. Therefore, one of the contributions of this paper is appropriately identifying children and creating a set of explanatory variables for study. For more information on this process, please see the data appendix.

Among the native-born children under study, 84 percent speak only English at home, with this proportion ranging from a high of 95 percent for White non-Hispanic children to a low of 38 percent among Hispanic children (Table 1). Spanish is by far the most prominent language after English for native-born Hispanic children (99 percent), but it is also the most frequent language among the U.S.-born White non-Hispanic (35 percent), Black (48 percent) and Indigenous (33 percent) children. Among the native-born Asian children, Chinese dominates (34 percent for Chinese and Cantonese combined), followed by Vietnamese, Korean and Tagalog.

Among the native-born White non-Hispanic children, the languages spoken after Spanish (35 percent) include German (10 percent), French (7 percent), Arabic (6 percent) and Amish (4 percent). While among the native-born Indigenous children, after Spanish comes Navaho (28 percent), followed by Aleut-Eskimo languages (14 percent), and then a variety of Native

[1] The percent of children who have Foreign Born parents with English-speaking ancestry is trivial (1% for both parents to 4% for mother or father). This signifies that an English-speaking ancestry does not drive any results.

American languages. While among native-born Blacks only 7 percent report speaking a language other than English, Spanish is dominant among foreign languages (48 percent), followed by French Creole (19 percent) and French (8 percent), with 8 percent speaking the African languages Kru, Ibo, Yoruka and Amharic.

Among the 1.5 generation children, only 28 percent speak only English at home (Table 2). For the children who speak a language other than English, Spanish dominates overall (55 percent), and among Hispanics (99 percent), while the modal language is Russian (19 percent) among white non-Hispanics, Tagalog (21 percent) for the Asian children, and French Creole (23 percent) for the Black children. Among the white non-Hispanic 1.5 generation (immigrant) children, after Russian come Arabic (13 percent), German (7 percent), French (6 percent) and Portuguese (5 percent). Overall, however, while a myriad of languages are to be found among bilingual children in the US, the overwhelming proportion of children have Spanish as their non-English language both among those born in the US and among the foreign born who immigrated as children.

A model of the determinants of destination language proficiency has been developed and used extensively for analyses of adult immigrant linguistic adjustment (Chiswick and Miller, 2007). This model is based on three conceptual variables: *Exposure* to the destination language, *Efficiency* in language acquisition, and *Economic incentive* for immigrant language acquisition. This approach is applied here for the determinants of bilingualism among children in the US. The emphasis, however, is on the variables related to exposure, as the efficiency and economic incentive effects so dominant in analyses for adult immigrants are assumed to play a minor role for childhood bilingualism.

The variables used in the statistical analysis are:

Dependent Variable:

Childspkhome: this binary dependent variable has a value of 1 if the child speaks a language other than English at home and 0 if he or she does not.

Explanatory Variables[2]

Female: A binary variable for the child's gender (female=1). This variable is expected to have a positive coefficient. As Lutz and Crist (2011) showed, particularly in Hispanic households, the mother is more likely to teach the foreign language to her daughters.

Foreign-born grandparent: A binary variable for whether a foreign-born grandparent is present in the household (Foreign born grandparent=1). This variable is expected to have a positive coefficient as children have greater exposure to a language other than English.

1+ Parent Prof: At least one parent speaks only English at home or speaks English "very well" or "well" (=1). This variable is expected to have a negative coefficient because greater parental English proficiency makes it more likely that English rather than another language is spoken in the home.

1+ Parent FB: At least one parent is foreign born (=1). This variable's coefficient is anticipated to be positive, as the presence of a foreign-born parent would increase the likelihood that the child is exposed to and hence speaks a foreign language in the home.

[2] For more details on the identification of family members and the construction of variables, see the data appendix.

Father (Mother) Educ: Parental years of schooling. As parental English proficiency increases with parental level of schooling, the more likely it is that the children of more educated parents are exposed to and speak only English.

Same Ancestry: The father and mother report the same ancestry (Ancestry=1). If the parents are of the same ancestry, they are more likely to speak a common language other than English at home, as will the child. Moreover, speaking the language may have a cultural preservation meaning for the household, which is more likely if the parents are of the same ancestry.

Years Since Migration (YSM): YSM is measured in years for the 1.5 generation. It is hypothesized, holding age constant, that the longer the child has lived in the U.S., the greater the degree of Americanization (exposure to English) and the lower the probability of the child being bilingual.

Child Order: This is a count variable for the order of birth of the child in the household. For example, if there are three children in the household, the oldest child is coded as 1, the middle child a 2, and the youngest child a 3. An only child is coded as 1in the child order variable. This variable is expected to have a negative coefficient as older children are more likely to speak a language other than English at home, due to their being an only child for at least a period of time, and hence having received greater parental attention.

Only Child: This is a dichotomous variable coded as 1 if the child is the only child in the household. This variable is expected to have a positive coefficient, due to the greater parental attention received. This variable permits an identification of the difference between being the oldest child in a multi-child family and being an only child.

Asian, *Hispanic, Indigenous, Black*: These are binary variables for race/ethnicity as defined by the Census Bureau. If the child is reported as Asian (not multiracial) in the race question, he/she is coded as 1 for *Asian*.[3] If the child is reported as being Hispanic, Latin, or of Spanish origin, he/she is coded as 1 for *Hispanic*. If the child is reported as American Indian or Alaska Native, he/she is coded as 1 for *Indigenous*. Similarly, if the child is reported as Black in the race question, he/she is coded as 1 for *Black*. We expect the coefficients on *Asian*, *Hispanic,* and *Indigenous* to be positive and significant, given the strong likelihood of a background which may include a foreign language.

Concentration: The percentage of those in the metropolitan area (if the child is living in a metropolitan area) who speak the same language as either the child, or the modal language in the household if the child speaks only English but other household members speak another language[4]. This variable is included in various forms, either continuously or dichotomously, with a threshold in robustness checks.

Table 3 presents the means (and standard deviations) for the sample used for the regression analysis. For some variables the household characteristics of the foreign-born children differ substantially from those of native-born children. For example, only 16% of native-born children speak a language other than English at home, while 72% of foreign-born children do. The proportions of the sample who are Asian or Hispanic are much greater for the foreign-born (18% and 42%, respectively), than for the native-born (3% and 15%), highlighting the

[3] "Asian Alone" is further narrowed by ancestry to reflect subgroups. See appendix for details.

[4] This variable is based on the definition of a Metropolitan Statistical Area (MSA) by the Census Bureau. See data appendix for more detail.

importance of origin for these groups.[5] The mean concentration ratio is 2.5 times higher for the foreign born children than it is for the native born (8 percent vs. 3 percent, respectively). Among those who live in an area with a non-zero concentration ratio for the language spoken at home, there is virtually no difference in the concentration ratio between foreign-born and native-born children (14% vs. 16%). Regarding household composition, the mean number of children in the sample is 1.5 per household, with 27% single-child households.

Interestingly, 22% of native-born children in the sample who speak another language at home do not have at least one foreign-born parent or a foreign-born grandparent (living with them), which suggests a strong desire to maintain the cultural heritage via language even across generations or that there is another source of their foreign language skills. Although 61% of those children speak Spanish, only 51% of them identify as Hispanic. The remaining children speak a variety of languages with German (8.1%), French (6.3%), Pennsylvania Dutch (4.3%), Yiddish (2.7%), Italian (2.2%), and Navaho (1.9%) being the next largest groups, with 12 percent speaking a large variety of other languages. These patterns may indeed be indicative of attempts at cultural preservation, but perhaps also of a foreign-language speaking caregiver.

Among the native-born children who speak a foreign language at home, 25% (or 19% if Spanish ancestries are grouped together) have parents of different ancestries, suggesting that in these families perhaps one parent may be instrumental in maintaining the language, even if it is

[5] Percent of Asian children who speak a language in addition to English, by parental ancestry

	NB	1.5 Gen
Neither Parent is of Asian Ancestry	1	3
Asian Father Only	1	1
Asian Mother Only	2	5
Asian Father and Mother	96	91
Total	**100**	**100**

Source: American Community Survey, IPUMS, 2005-2011

not spoken by their spouse. Alternatively, the parents may speak the same language with different ancestral backgrounds. This is especially true for a language used across different ethnicities and countries of origin, such as Spanish.

IV. Regression Analysis

The analysis in Table 4, Column I is an Ordinary Least Squares (OLS) regression of *childspkshome* on several explanatory variables, including gender, parental language proficiency, parental nativity, parental education, parental ancestry, child's race/ethnicity, and child's birth order in the household.

The regression coefficients are all statistically significant (large sample sizes) and the regression has high explanatory power ($R^2 = 0.57$). As anticipated by our model, and consistent with previous literature, the variables *female, foreign born grandparent, 1+ foreign born parent, only child* and *same ancestry,* all have a positive effect on whether the child speaks another language at home. Greater *parental proficiency* in English, greater *parental education*, and *child order* have the expected negative impact on bilingualism.

Adding the enclave measure, *concentration*, to the equation (compare Table 4, columns 1 and 2) has little effect on the magnitudes of the other variables, with the exception of *Hispanic*. The magnitude of being Hispanic drops substantially from 0.28 to 0.20 in response to the inclusion of the concentration measure, partially due to a substantial correlation between being Hispanic and a high concentration ratio (0.63). Hispanic children are more likely to speak a language in addition to English (nearly exclusively Spanish) at home, and live in areas with a large concentration of Spanish speakers.

Since all of the concentration ratios greater than 0.16 are for the Spanish language, it is vital to run regressions separately by ethnic group. Table 4 (columns 3 to 6) analyze the effects by gender and Hispanic/Non-Hispanic ethnicity, and Table 5 adds analyses for 6 other ethnic groups: Asian, Chinese, Korean, Cuban, Dominican, and Mexican.

Tables 4 and 5 demonstrate that for many of the explanatory variables, the magnitudes and signs of the regression coefficients do not vary substantially across groups, regardless of gender and ethnicity. What then are the most significant factors? They appear to be parental English language proficiency, parental nativity, parents of the same ancestry, and of course, the linguistic concentration measure.

The effect of having at least one parent proficient in English lowers the probability that a child will be bilingual. However, this coefficient has a wide range, from a low of -0.06 for Mexican children (and similarly -0.07 to -0.09 for Cubans, Dominicans, and Hispanics overall) to -0.38 for the Non-Hispanic groups. This suggests that even controlling for the concentration variable, parental English language proficiency is a much more important determinant of bilingualism among children for non-Spanish speaking groups.

The effects of having at least one foreign-born parent are positive and significant and have a wide range, but the pattern is less clear. The magnitude ranges from 0.24 for Cuban children to 0.53 for Mexican children. The coefficients do not appear to be bunched by ethnicity overall or language group.

Interestingly, though not surprisingly, common parental ancestry has one of the biggest impacts on whether a child speaks a language other than English at home. Since *same ancestry* is defined as a couple having the same ancestry code, this effect may actually be a lower bound for

the true language effect, particularly for Hispanics. For example, someone of Argentinian ancestry may marry someone of Mexican ancestry and have different ancestry codes, but raise Spanish-speaking children[6]. As there is only one code to identify those of Korean ancestry (and thus likely most of the Korean speakers), the particularly high coefficient on "same parental ancestry" for those of Korean origin (0.49) lends some support to the idea of the lower bound on the same ancestry coefficient overall. There will be a particular downward bias for children whose parents report different ancestries, but speak the same non-English language in the household. This will be discussed further in the robustness section.

The regression coefficient of the linguistic *concentration* ratio varies substantially across ancestry groups and subgroups among Hispanics and non-Hispanics. For example, for Asians overall, the concentration coefficient is only 0.25, but for Chinese children and Korean children, the magnitudes are 1.32 and 1.15, respectively. Similarly, for Mexican children the concentration coefficient is 0.30 but for Dominican and Cuban children, the effects are 0.53 and 0.61, respectively. When New York City is excluded, the effect of the concentration coefficient becomes stronger for many of the smaller groups (see table in Appendix). These results suggest that it is not the specific *language* impacting the results, but a clear country of origin-specific community effect.

Analyzing the language concentration variable demonstrates that the effect on bilingualism of living in a linguistic concentration area is considerably more important for non-Hispanic groups than for Hispanic groups. Given the prevalence of the Spanish language outside the Hispanic enclaves, children of Hispanic origin can be exposed to the language even if they do

[6] To reduce this tendency, the 6 codes relating to Mexican ancestry and 2 for Chinese ancestry have been combined.

not live in the enclaves. Because of the virtual absence of exposure to other foreign languages outside their enclaves, children from these language groups are that much more likely to become bilingual if they live inside their linguistic enclave. Accordingly, the coefficient on *Concentration* is 1.2 for non-Hispanic groups, thrice that of 0.4 for Hispanic groups.

As seen in Table 6, the frequency distribution of the concentration measure varies widely by racial/ethnic group. The distribution appears to be U-shaped for Spanish-speaking groups. There is a clear decline in the percent of children in the higher concentration index categories for non-Hispanic groups overall and East-Asian groups. While not wholly surprising, the differences in these patterns highlight the importance of examining ethnic groups separately in addition to together, and help to explain the large variance in the magnitude and importance of the concentration coefficient. Although it may be hypothesized that many of these groups are concentrated in New York City, excluding all boroughs does not significantly impact the distribution of the linguistic concentration measure (see table in Appendix).

Although gender is discussed as a very important determinant of bilingualism in the literature, the effect is not very large. Girls *are* more likely than boys to speak a foreign language, but only with magnitudes ranging from 1 percentage point (overall) to 4 percentage points (Korean).

The effect of a foreign-born grandparent in the household is surprisingly strong for some groups, but insignificant for others. Though it is insignificant for Hispanics and all Spanish-speaking groups, the coefficient is 0.05 for Non-Hispanic groups and 0.07 for Chinese children. While there is a strong incentive for children (whether self-motivated or encouraged by the parents) to learn the language to communicate with their grandparents, the presence of a foreign-

born grandparent in the household may also indicate a recent migration, or in general, close generational and cultural ties which may contribute to an environment in which the child is encouraged to speak the origin language at home. Perhaps this grandparent effect is not strong for Hispanic groups due to other opportunities for Spanish language acquisition in the ethnic community, which may not exists for other language groups.

As expected, ceteris parabis, older children are more likely to speak a foreign language in the home, due to perhaps less assimilation or more time input by the parents, as he/she is at least for some time the only child.[7] The effect of *child order* (Table 5) has the expected negative sign for all groups, though it is not always significant. The negative effect is strongest for Korean and Dominican children (3% each), with 2% for Hispanic groups overall, but insignificant for non-Hispanics. However, the pattern is different when the child has no siblings. Though in Table 4 it appears that an *only child* is weakly more likely to speak a foreign language at home, we can see in Table 5 that this effect is stronger for Asian groups (3% for Korean children and 5% for Chinese).

Although more educated parents are more likely to speak English at home, this is not a major determinant of childhood language usage. Consistent with the findings of Biedinger et al. (2014), the greater the education of the parent, the less likely the child is to speak the origin language. While parental education has the predicted negative impact on the child's bilingualism,

[7] Among the foreign-born children in the sample, it would be expected that those who arrived at a younger age would be more likely to be monolingual English speakers. The mean age of migration is 4.5 years for foreign-born children, age 5 to 18 in the ACS: that is, a mean age of migration of 2.9 years for those who don't speak a foreign language at home, and 5.1 years for those that do.

the magnitude is very low, ranging from insignificant to -0.013 (Asian) for both mother and father.

The analysis in Table 7 considers whether the model differs by the actual nativity of the child. Recall that the sample includes foreign-born children who came to the US before age 14 (1.5 generation). All else constant, the 1.5 generation is 16 percent (t-stat=115.6) more likely to speak a foreign language at home than the native-born. Table 7, Column 1 repeats the previous results of Table 2, Column 1 for comparison. Due to the relative sizes of the groups, the results for "All" are driven by the US-born children (95% of the sample). However, there are stark differences between the groups. For example, the presence of a foreign-born grandparent in the household is important for native-born children (0.05), but insignificant for 1.5 generation children. A child born in a foreign country (1.5 generation) has been exposed to the language in the origin, with opportunity to develop these linguistic skills. Conversely, a native-born child may have limited or no opportunity to absorb a foreign language, other than in the household. A foreign-born grandparent, especially one who may provide child care, either in the absence of or in addition to the parents, would be instrumental in helping a child develop those origin linguistic skills. Similarly, parental proficiency in English (negative effect of greater proficiency on bilingualism) matters less for the 1.5 generation than the native-born, but having at least one foreign-born parent matters more for speaking another language for the 1.5 generation (positive effect on bilingualism). Mother's education has a significant negative effect on bilingualism for both nativity groups, while father's education only has a significant negative effect for the native-born children. If a 1.5 generation child is an only child, he/she is 2% more likely to speak a foreign language in the home; however, there is no effect of *only child* among native-born

children. As expected, *same ancestry* among the parents is very important in promoting bilingualism (again: likely a lower bound), though much more so for 1.5 generation children.

The most striking difference between the two groups of children is in the *concentration* variable. Living in a metropolitan area with a high linguistic concentration has a *much* smaller (one-fifth in size) effect for the 1.5 generation (Table 7). These striking results imply that exposure to the language of the origin country has a large impact on language retention and usage at home, such that the enclave may be helpful, but is not nearly as important for the 1.5 generation as for native-born children.

Table 8 reports the analyses by household structure – two parent, father only, and mother only households. A father-only household vs. a mother-only household does not yield very different results in terms of childhood bilingual acquisition. The presence of a foreign-born grandparent appears to matter more in a father-only household, perhaps because the grandparent is more likely to serve as a second parent substitute in this situation.

In Table 9, four additional variables are added to the basic model: the child's age, whether the state of residence borders Mexico or is Florida, and whether the child is Black or Indigenous, with a fifth variable, years since migration, included for the 1.5 generation. The children in the sample are all age 5 to 18 years. Older children are more likely to report being bilingual and the effect of age is larger for the 1.5 generation. When age is included in the analysis, the effect of *child order* declines, and becomes statistically insignificant for the 1.5 generation, though the effect of *only child* is unchanged.

Other variables being the same, *Black* children are less likely to be bilingual, 7 percent less likely in the pooled sample. Native-born children of *Indigenous* origins are more likely to be

bilingual, as the concentration measure does not reflect that many of them live in indigenous communities outside of metropolitan areas on or near Indian reservations or Alaska. There were very few 1.5 generation Indigenous children.

Living in a *Border State* is associated with a lower propensity for bilingualism overall, but its coefficient is very small and not statistically significant for the 1.5 generation. The lower probability of bilingualism from living in a Spanish language intensive border state may arise from the exposure effect being captured by the concentration measure.

Finally, among the 1.5 generation, the effect of a longer time in the U.S. (greater YSM), that is, arriving at a younger age since current age is held constant, lowers the probability of being bilingual. This is to be expected as there has been a longer period of being exposed to English.

Robustness

A number of statistical tests are performed to ascertain the robustness of the analysis. The first robustness check comes from a comparison between parental proficiency variables. Paternal and maternal proficiency in English are highly correlated (R=0.86). Similarly, there was a high degree of collinearity between paternal and maternal age at migration (R=0.79). These correlations stem from the correlation of paternal and maternal foreign birth (R=0.74). The regressions from Table 5 were recomputed including more detail on parental proficiency and nativity. Table 10 demonstrates that it makes little difference whether the included variables for parental English proficiency or parents foreign born are "1+ parent", father only, or mother only.

Table 10 also shows that in almost all cases, the coefficients of the parental proficiency and nativity variables are of similar magnitude and always of the same sign. While the maternal

language proficiency coefficients are slightly larger, the maternal foreign birth coefficients are slightly smaller for all three groups. The parental age at migration coefficients are identical. These results justify using "at least one parent foreign born" and "at least one parent proficient".

Another variable for which robustness checks were conducted was the *concentration* measure. To check whether the concentration variable sufficiently controlled for an enclave effect, it was introduced as a set of dichotomous variables at various thresholds, such as 1%, 5%, 10%, etc. Bilingualism increased with the concentration measure, although any concentration above 16% is only for the Spanish language. We tested for a nonlinear effect of the concentration measure by adding a squared term, but the linear and squared terms were highly collinear and no nonlinear effect was found. Additionally, there were no noteworthy changes in the other coefficients in any of the above specifications.

Statistical methodology was also tested. Since the dependent variable is binary, natural choices for statistical estimation are OLS, probit and logit models. The comparison of these models is in Table 11. All coefficients have the same signs and display similar magnitudes relative to each other across the three statistical models. Although it is difficult to interpret the coefficients in probit and logit models (unless marginal effects are computed), this analysis reinforces the validity of the specification adopted.

It is also useful to test whether the regression results are sensitive to the coding of the Spanish-origin ancestry groups. Therefore, all ancestry codes of predominantly Spanish speaking countries of Latin America and the Caribbean were recoded to be one ancestry. As seen from a comparison of Table 4 and Table 12, the magnitude of the ancestry coefficient (0.053) is, as expected, unchanged for the non-Hispanic groups. Note, however, that there was a large increase

in the ancestry coefficient for Hispanics (from 0.18 to 0.24). Pooling ancestries for other common languages might lead to a similar increase in the effect of the linguistic concentration index on bilingualism for groups that speak those languages. However, while Spanish-speakers constitute 12% of the sample (and 65% of children who speak a language other than English at home), there is no other language group that is dominant, so the effect of any additional grouping of ancestries by language is likely to be marginal.

V. Summary and Conclusion

This paper uses the American Community Survey (2005-2011) as assembled by IPUMS to analyze the determinants of bilingualism for native-born and 1.5 generation (immigrated before age 14) children. The model focuses on variables associated with *exposure* to the language other than English as the primary determinants of bilingualism among children. Parental proficiency in English, parental foreign birthplace, residence in a linguistic enclave, shared parental ancestry, and race/ethnicity are found to be the most important determinants of child bilingualism across all groups. Overall, having at least one parent proficient in English decreases the probability that a child will speak a language other than English by 19%, while having at least one parent born abroad increases the probability by 39%; a 1 standard deviation increase in parental foreign birth also leads to a 0.42 standard deviation increase in child bilingualism, a very strong, but not unanticipated, effect.

Additionally, Asian and Hispanic children are 19% and 20%, respectively, more likely to speak a language other than English at home than are comparable white non-Hispanic children, even when the enclave effect is held constant. U.S.-born Indigenous children are more likely to be bilingual, but Black children are less likely to be bilingual, than white non-Hispanic children.

A 1 standard deviation increase in the linguistic concentration index leads to a 0.15 standard deviation increase in the probability that a child will speak a language other than English. Among the 1.5 generation, those who immigrated at a younger age are less likely to be bilingual. The significant negative effect of parental education and the positive effect of a foreign-born grandparent living in the household appear to be largely overshadowed by the other variables. Older children in a multi-child family and only children are more likely to be bilingual.

This study also demonstrates the importance of the analysis of bilingualism separately by ethnic group. For example, despite high coefficients on *concentration* for non-Hispanic groups, a standard deviation analysis comparing magnitudes demonstrates that although the concentration index is very important for the native-born children and Spanish-speaking groups, it is much less important for the 1.5 generation and Asian groups. When looking at regressions computed separately by ethnicity, shared parental ancestry becomes very important – even more important than parental nativity for Koreans and Cubans.

Although these results provide interesting insights, they also raise new questions. There may be important unmeasured determinants of bilingualism. Cultural characteristics may determine childhood language acquisition, such as the importance to the parents of the language in maintaining an origin culture. It's also likely that families who consider return migration in the future will not only speak the origin language in the household, but may also make a greater effort to teach the language to their children. This may be particularly relevant for immigrants from Mexico, as they have a greater propensity for return migration. Unfortunately, data are unavailable on the level of proficiency in the foreign language, which precludes an analysis of the determinants of both the proficiency in that language and whether this detracts from their

English language skills. Moreover, it would be useful to know who among the bilingual children in these data will retain their bilingualism as adults, and the extent, if at all, to which native born and 1.5 generation monolingual English-speaking children may eventually become bilingual adults.

References

Bak, Thomas H., Jack J. Nissan, Michael M. Allerhand, and Ian J. Dreary (2014), "Does Bilingualism Influence Cognitive Aging?" *Annals of Neurology*. 75(6), pp. 959-963.

Biedinger, Nicole, Birgit Becker and Oliver Klein (2014), "Turkish-language Ability of Children of Immigrants in Germany: Which Contexts of Exposure Influence Preschool Children's Acquisition of their Heritage Language?" Paper presented at the "Turkish Migration Conference: Comparative Perspectives and Continuities," Regent's University, London, June 2014.

Chiswick, Barry R., Yew Liang Lee, and Paul W. Miller (2005), "Parents and Children Talk: English Language Proficiency Within Immigrant Families" *Review of Economics of the Household* 3(3), pp. 243-268

Chiswick, Barry R. and Miller, Paul W., (2007). *The Economics of Language: International Analyses*, London: Routledge.

Chiswick, Barry R. and Miller, Paul W. (2013), "Does Bilingualism Among the Native Born Pay", Working Paper

Feliciano, Cynthia (2001), "The Benefits of Biculturalism: Exposure to Immigrant Culture and Dropping out of School among Asian and Latino Youths", *Social Science Quarterly*, 82(4).

Feng, Li. Yunwei Gai and Xiaoning Chen (2014) "Family Learning Environment and Early Literacy: A comparison of Bilingual and Monolingual Children" *Economics of Education Review*, 30, pp. 110-130.

Lutz, Amy and Crist, Stephanie (2011), "Why do Bilingual Boys get Better Grades in English-only America? The Impacts of Gender, Language and Family Interaction on Academic Achievement of Latino/a Children of Immigrants" *Ethnic and Racial Studies*, 32(2), pp.

346-368.

Paradis, Johanne (2007), *Early Bilingual and Multilingual Acquisition* in Peter Auer and Li Wei (eds.) Handbook of Multilingualism and Multilingual Communication, Vol. 5, Berlin Germany, pp. 15-44

Pendakur, Krishna and Pendakur, Ravi, "Language as Both Human Capital and Ethnicity", *International Migration Review*, 36(1), pp. 147-177

Portes, Alejandro and Rivas, Alejandro (2011), "The Adaptation of Migrant Children", *Immigrant Children,* 21(1).

Saiz, Albert and Zoido, Elena (2012), "The Returns to Speaking a Second Language", Working Paper, Federal Reserve Bank of Philadelphia, Philadelphia, PA, USA.

Tables

Table 1: Languages Spoken by Native-Born Children at Home (Percent)

	All	White, non-Hispanic	Hispanic	Asian	Black	Indigenous
% that speak only English	84	95	38	41	93	77
For those who speak a language other than English (percent)						
Top 5 languages	Spanish 68	Spanish 35	Spanish 99	Chinese 27	Spanish 48	Spanish 33
	Chinese 3	German 10	French 0.1	Vietn. 22	French Creole 19	Navaho 28
	Vietn. 2	French 7	Portuguese 0.1	Korean 12	French 8	Aleut-Eskimo Languages 14
	German 2	Arabic 6	German 0.1	Tagalog 7	Kru, Ibo, Yoruba 5	Other N. American Indian Languages 9
	French 2	Pennsylvania Dutch 4	Japanese 0	Cantonese 7	Amharic 3	Keres 3
All other languages	Other 23	Other 38	Other 0.7	Other 25	Other 17	Other (a) 13
Total	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**
N	**409,286**	**88,016**	**243,486**	**42,503**	**14,430**	**5,780**

(a) The 13 percent speaking other languages included 8 percent for other American Indian languages, 1.5 percent other European languages, 1.5 percent South Asian languages, and 1.6 percent other languages

Source: American Community Survey, IPUMS 2005-2011

Table 2: Languages Spoken by Foreign-Born (1.5 Generation) Children at Home (Percent)

	All	**White, non-Hispanic**	**Hispanic**	**Asian**	**Black**
% that speak only English	28	52	7	31	50
For those who speak a language other than English (Percent)					
Top 5 languages	Spanish 55	Russian 19	Spanish 99	Tagalog 21	French Creole 23
	Tagalog 4	Arabic 13	Portuguese 0.1	Chinese 18	Spanish 14
	Russian 3	German 7	German 0.1	Korean 17	Kru, Ibo, Yoruba 13
	Chinese 3	French 6	French 0.1	Vitenm. 14	French 9
	Korean 3	Portuguese 5	South/Central American Indian Languages 0.1	Japanese 6	Cushite 9
All other languages	Other 32	Other 50	Other 0.6	Other 24	Other 32
Total	**100%**	**100%**	**100%**	**100%**	**100%**
N	**105,114**	**17,265**	**56,735**	**17,892**	**5,464**

Source: American Community Survey, IPUMS 2005-2011

Table 3: Descriptive Statistics: Means of Variables (Standard Deviations), Children, Age 5 to 18 Living at Home

Variable	Full Sample	Foreign Born[(1)]	Native Born
Child Speaks a Language Other Than or in Addition to English at Home	0.190 (0.392)	0.724 (0.447)	0.160 (0.366)
Female Child	0.486 (0.500)	0.491 (0.500)	0.486 (0.500)
At least 1 Parent Proficient in English	0.961 (0.193)	0.755 (0.430)	0.973 (0.162)
At least 1 Parent Foreign Born	0.241 (0.428)	0.859 (0.348)	0.206 (0.405)
Father Years of Education	13.456 (3.201)	12.616 (4.664)	13.504 (3.091)
Mother Years of Education	13.521 (2.993)	12.203 (4.494)	13.595 (2.867)
Same Ancestry for Both Parents	0.495 (0.500)	0.745 (0.436)	0.481 (0.500)
Child Birth Order (oldest=1)	1.569 (0.802)	1.463 (0.749)	1.575 (0.804)
Only Child	0.265 (0.442)	0.278 (0.448)	0.265 (0.441)
Foreign-Born Grandparent Present	0.032 (0.175)	0.100 (0.300)	0.028 (0.165)
Asian	0.036 (0.187)	0.180 (0.384)	0.028 (0.166)
Hispanic	0.167 (0.373)	0.419 (0.493)	0.153 (0.360)
Black	0.078 (0.269)	0.075 (0.264)	0.079 (0.269)
Indigenous	0.009 (0.097)	0.005 (0.073)	0.010 (0.098)
Concentration	0.033 (0.100)	0.081 (0.141)	0.030 (0.096)
N	**2,711,065**	**145,175**	**2,565,890**

(1) Migrated to the US below age 14.
Standard deviations in parentheses.
Source: American Community Survey, IPUMS, 2005-2011

Table 4: Regression Analysis of Child Speaking a Language Other than English at Home

	Overall	Overall	Male	Female	Hispanic	Non-Hispanic
Female	0.010**	0.010**			0.020**	0.008**
	(27.27)	(26.88)			(17.92)	(21.67)
Foreign Born	0.051**	0.033**	0.030**	0.036**	0.001	0.051**
Grandparent	(24.07)	(14.92)	(11.81)	(12.36)	(0.26)	(14.38)
1+ Parent Prof.	-0.204**	-0.190**	-0.202**	-0.176**	-0.076**	-0.383**
	(151.71)	(136.44)	(131.36)	(97.24)	(60.44)	(103.86)
1+ Parent FB	0.405**	0.388**	0.382**	0.394**	0.492**	0.327**
	(338.45)	(319.89)	(266.32)	(228.96)	(180.79)	(243.20)
Father Educ	-0.003**	-0.002**	-0.003**	-0.002**	-0.005**	0.001**
	(21.68)	(19.97)	(17.94)	(15.58)	(24.45)	(7.34)
Mother Educ	-0.007**	-0.007**	-0.007**	-0.007**	-0.008**	-0.005**
	(58.23)	(56.94)	(48.60)	(46.37)	(43.28)	(33.96)
Same Ancestry	0.077**	0.072**	0.071**	0.072**	0.181**	0.053**
	(141.88)	(130.44)	(115.80)	(96.58)	(77.45)	(103.57)
Child Order	-0.004**	-0.004**	-0.002**	-0.005**	-0.017**	0.000
	(10.62)	(11.04)	(5.91)	(11.70)	(20.01)	(0.17)
Only Child	0.001*	0.002**	0.002**	0.002	0.002	0.006**
	(2.13)	(3.07)	(3.04)	(1.86)	(1.16)	(10.31)
Asian	0.169**	0.185**	0.183**	0.186**	(a)	0.210**
	(83.77)	(89.58)	(67.92)	(73.13)		(96.50)
Hispanic	0.282**	0.200**	0.199**	0.202**	(a)	(a)
	(227.06)	(156.26)	(113.68)	(119.72)		
Concentration	(a)	0.586**	0.564**	0.608**	0.397**	1.197**
		(131.40)	(96.81)	(106.39)	(77.49)	(65.77)
_cons	0.327**	0.312**	0.328**	0.307**	0.363**	0.442**
	(189.64)	(171.24)	(153.24)	(121.07)	(93.82)	(118.23)
R^2	0.57	0.58	0.58	0.58	0.49	0.36
N	**2,711,065**	**2,711,065**	**1,393,932**	**1,317,133**	**453,928**	**2,257,137**

Note: t-statistics in parentheses, (a) designates variable not entered
* significant at 5% level, ** significant at 1% level
Source: American Community Survey, IPUMS 2005-2011

Table 5: Regression Analysis by Ethnicity
Whether a child speaks a language at home other than English

	Asian	Chinese	Korean	Cuban	Dominican	Mexican
Female	0.025**	0.021**	0.041**	0.026**	0.020*	0.017**
	(8.97)	(4.38)	(6.01)	(3.26)	(2.44)	(12.72)
Foreign Born	-0.005	0.070**	0.015	0.010	-0.004	-0.002
Grandparent	(0.76)	(7.78)	(0.71)	(0.72)	(0.23)	(0.55)
1+ Parent Prof.	-0.226**	-0.154**	-0.164**	-0.085**	-0.066**	-0.064**
	(43.90)	(16.59)	(16.38)	(8.10)	(8.63)	(45.04)
1+ Parent FB	0.478**	0.440**	0.344**	0.240**	0.366**	0.526**
	(86.00)	(45.94)	(25.86)	(16.25)	(11.30)	(137.74)
Father Educ	0.004**	-0.002	0.006**	-0.008**	-0.006**	-0.005**
	(6.16)	(1.70)	(2.73)	(5.11)	(3.52)	(16.70)
Mother Educ	-0.013**	-0.007**	-0.011**	-0.002	-0.007**	-0.008**
	(21.85)	(7.13)	(4.78)	(1.16)	(3.98)	(34.98)
Same Ancestry	0.181**	0.292**	0.490**	0.231**	0.168**	0.190**
	(33.73)	(31.53)	(40.59)	(17.10)	(10.82)	(52.77)
Child Order	-0.013**	-0.026**	-0.033**	-0.008	-0.032**	-0.015**
	(7.67)	(7.37)	(4.87)	(1.14)	(5.85)	(15.65)
Only Child	0.041**	0.048**	0.033**	0.027**	-0.018	-0.006**
	(7.60)	(8.55)	(3.88)	(2.49)	(1.77)	(2.78)
Concentration	0.250**	1.321**	1.148*	0.605**	0.534**	0.300**
	(2.82)	(8.23)	(2.48)	(27.06)	(10.42)	(48.16)
_cons	0.360**	0.274**	0.187**	0.396**	0.564**	0.329**
	(31.33)	(17.86)	(4.67)	(11.09)	(14.38)	(63.78)
R^2	0.21	0.37	0.46	0.43	0.25	0.50
N	**98,452**	**31,437**	**14,113**	**10,669**	**6,942**	**274,815**

Note: t-statistics in parentheses. Asian includes persons of Chinese, Korean ancestry, East Asian, and Southeast Asian ancestries.
*significant at 5% level, ** significant at 1% level
Source: American Community Survey, IPUMS, 2005-2011

Table 6: Frequency Distribution of Concentration by Group: Percent of Children in Each Race/Ethnicity group in the Concentration Categories

Concentration	Asian	Chinese	Korean	Cuban	Dominican	Mexican	Hispanic	Non-Hispanic
0%	30.45	25.41	30.22	19.07	7.46	23.79	27.05	88.94
<0.5%	18.27	19.94	25.75	0.29	0.22	0.15	0.39	3.59
0.5%-1%	16.43	20.67	12.68	0.15	0.19	0.05	0.16	1.66
1%-2%	13.09	10.20	21.26	0.62	0.79	0.41	0.57	1.64
2%-3%	8.86	12.72	5.70	1.48	2.07	0.97	1.36	0.98
3%-4%	4.28	2.73	3.54	0.78	0.82	0.98	1.05	0.51
4%-5%	3.30	4.83	0.19	1.96	3.44	1.59	1.96	0.43
5%-10%	4.74	2.76	0.17	10.26	21.22	6.25	8.97	0.94
10+ %	0.58	0.75	0.49	65.39	63.79	65.80	58.67	1.30
Total	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**

Source: American Community Survey, IPUMS, 2005-2011

Table 7: Regression Analysis by Child Nativity

Whether a child speaks a language at home other than English

	All	Native Born	1.5 Generation
Female	0.010**	0.011**	0.006**
	(26.88)	(26.72)	(3.07)
Foreign Born Grandparent	0.033**	0.045**	-0.002
	(14.92)	(19.63)	(0.45)
1+ Parent Prof.	-0.190**	-0.196**	-0.086**
	(136.44)	(127.82)	(34.09)
1+ Parent FB	0.388**	0.341**	0.583**
	(319.89)	(253.49)	(141.31)
Father Educ	-0.002**	-0.003**	0.000
	(19.97)	(26.80)	(1.08)
Mother Educ	-0.007**	-0.007**	-0.003**
	(56.94)	(55.05)	(7.26)
Same ancestry	0.072**	0.062**	0.154**
	(130.44)	(115.79)	(40.82)
Child Order	-0.004**	-0.001**	-0.005**
	(11.04)	(3.48)	(3.09)
Only Child	0.002**	0.001	0.020**
	(3.07)	(1.22)	(8.27)
Asian	0.185**	0.195**	0.094**
	(89.58)	(72.20)	(27.64)
Hispanic	0.200**	0.201**	0.192**
	(156.26)	(151.27)	(55.14)
Concentration	0.586**	0.672**	0.138**
	(131.40)	(136.26)	(19.17)
_cons	0.312**	0.334**	0.099**
	(171.24)	(165.58)	(16.32)
R^2	0.58	0.55	0.45
N	**2,711,065**	**2,565,890**	**145,175**

Note: t-statistics in parentheses

*significant at 5% level, ** significant at 1% level

Source: American Community Survey, IPUMS, 2005-2011

Table 8: Regression Analysis by Household Structure
Whether a child speaks a language at home other than English

	Both Parents	Father Only	Mother Only
Female	0.010**	0.012**	0.011**
	(26.88)	(8.39)	(17.25)
Foreign Born Grandparent	0.033**	0.152**	0.062**
	(14.92)	(16.77)	(12.72)
1+ Parent Prof.	-0.190**	(a)	(a)
	(136.44)		
1+ Parent FB	0.388**	(a)	(a)
	(319.89)		
Father Educ	-0.002**	-0.004**	(a)
	(19.97)	(11.20)	
Mother Educ	-0.007**	(a)	-0.004**
	(56.94)		(17.63)
Same Ancestry	0.072**	(a)	(a)
	(130.44)		
Child Order	-0.004**	-0.007**	-0.007**
	(13.89)	(6.68)	(14.46)
Only Child	0.002**	0.001	0.002*
	(3.07)	(0.91)	(2.12)
Asian	0.185**	0.221**	0.212**
	(89.58)	(26.24)	(29.06)
Hispanic	0.200**	0.234**	0.212**
	(156.26)	(56.16)	(101.30)
Concentration	0.586**	0.846**	0.918**
	(131.40)	(43.96)	(93.91)
Father Lang Prof	(a)	-0.193**	(a)
		(33.70)	
Father Foreign Born	(a)	0.362**	(a)
		(70.71)	
Mother Lang Prof	(a)	(a)	-0.224**
			(56.31)
Mother Foreign Born	(a)	(a)	0.339**
			(119.80)
_cons	0.312**	0.272**	0.295**
	(171.24)	(41.27)	(59.68)
R^2	0.58	0.60	0.52
N	**2,711,065**	**201,132**	**761,859**

Note: t-statistics in parentheses
(a) variable not entered
*significant at 5% level, ** significant at 1% level
Source: American Community Survey, IPUMS, 2005-2011

Table 9: Regression Analysis with additional variables

Whether a child speaks a language at home other than English

	Overall	Overall 2	1.5G	1.5G 2
Female	0.010**	0.010**	0.006**	0.007**
	(26.88)	(27.75)	(3.07)	(3.54)
Foreign Born Grandparent	0.033**	0.035**	-0.002	0.003
	(14.92)	(15.83)	(0.45)	(0.72)
1+ Parent Prof.	-0.190**	-0.185**	-0.086**	-0.072**
	(136.44)	(131.15)	(34.09)	(27.65)
1+ Parent FB	0.388**	0.392**	0.583**	0.556**
	(319.89)	(322.16)	(141.31)	(131.73)
Father Educ	-0.002**	-0.003**	0.000	0.000
	(19.97)	(21.54)	(1.08)	(0.41)
Mother Educ	-0.007**	-0.007**	-0.003**	-0.003**
	(56.94)	(56.99)	(7.26)	(9.04)
Same Ancestry	0.072**	0.078**	0.154**	0.155**
	(130.44)	(133.73)	(40.82)	(41.80)
Child Order	-0.004**	-0.001**	-0.005**	0.002
	(13.89)	(3.50)	(3.09)	(1.12)
Only Child	0.002**	0.004**	0.020**	0.020**
	(3.07)	(7.22)	(8.27)	(8.26)
Asian	0.185**	0.180**	0.094**	0.065**
	(89.58)	(87.37)	(27.64)	(20.92)
Hispanic	0.200**	0.198**	0.192**	0.168**
	(156.26)	(155.07)	(55.14)	(49.91)
Concentration	0.586**	0.612**	0.138**	0.141**
	(131.40)	(132.13)	(19.17)	(19.18)
Age	(a)	0.001**	(a)	0.010**
		(17.44)		(26.51)
Border State	(a)	-0.026**	(a)	-0.004
		(37.27)		(1.61)
Indigenous	(a)	0.066**	(a)	(a)
		(17.85)		
Black	(a)	-0.065**	(a)	-0.150**
		(73.39)		(21.59)
Years Since Migration	(a)	(a)	(a)	-0.011**
				(30.71)
Constant	0.312**	0.302**	0.099**	0.104**
	(171.24)	(148.10)	(16.32)	(12.11)
R^2	0.58	0.58	0.45	0.46
N	**2,711,065**	**2,711,065**	**145,175**	**145,175**

Note: t-statistics in parentheses

(a) Variable not entered

*significant at 5% level, ** significant at 1% level

Source: American Community Survey, IPUMS, 2005-2011

Table 10: Selected coefficients on variables for Parent Proficiency, Foreign Birth, and Age at Migration on whether a child speaks a language other than English at home

Coefficients	Overall	Non-Hispanic	Hispanic
1+ Parent Proficient	-0.190	-0.383	-0.076
Father Proficient	-0.188	-0.379	-0.096
Mother Proficient	-0.208	-0.405	-0.101
1+ Parent FB	0.388	0.327	0.492
Father FB	0.412	0.372	0.450
Mother FB	0.386	0.352	0.431
Father age at migration	0.015	0.015	0.015
Mother age at migration	0.015	0.015	0.015

Note: All coefficients significant at 1% level
Source: American Community Survey, IPUMS, 2005-2011

Table 11: Comparison of OLS, Probit and Logit Regressions on Whether a child speaks a language at home other than English

	OLS	Probit	Logit
Female	0.010**	0.082**	0.158**
	(26.88)	(30.40)	(30.68)
Foreign Born	0.033**	0.043**	0.058*
Grandparent	(14.92)	(4.88)	(3.79)
1+ Parent Prof.	-0.190**	-1.094**	-2.168**
	(136.44)	(68.55)	(61.19)
1+ Parent FB	0.388**	1.533**	2.817**
	(319.89)	(378.67)	(356.57)
Father Educ	-0.002**	-0.008**	-0.010**
	(19.97)	(10.68)	(7.12)
Mother Educ	-0.007**	-0.038**	-0.071**
	(56.94)	(48.80)	(47.25)
Same Ancestry	0.072**	0.455**	0.971**
	(130.44)	(105.29)	(113.96)
Child Order	-0.004**	-0.029**	-0.058**
	(11.04)	(11.13)	(11.55)
Only Child	0.002**	0.034**	0.082**
	(3.07)	(7.99)	(9.89)
Asian	0.185**	0.550**	0.889**
	(89.58)	(84.09)	(78.03)
Hispanic	0.200**	0.729**	1.311**
	(156.26)	(128.79)	(119.80)
Concentration	0.586**	2.629**	4.757**
	(131.40)	(98.55)	(90.86)
_cons	0.312**	-0.421**	-0.718**
	(171.24)	(22.39)	(17.63)
N	**2,711,065**	**2,711,065**	**2,711,065**

Note: t-stats in parentheses
* significant at 5% level, ** significant at 1% level
Source: American Community Survey, IPUMS 2005-2011

Table 12: Regression Analysis with all Spanish ancestry coded together
Whether a child speaks a language at home other than English

	Overall	Overall	Male	Female	Hispanic	Non-Hispanic
Female	0.010**	0.010**	(a)	(a)	0.020**	0.008**
	(27.16)	(26.79)			(18.06)	(21.67)
Foreign Born	0.051**	0.033**	0.030**	0.036**	0.002	0.051**
Grandparent	(24.06)	(15.04)	(11.91)	(12.44)	(0.65)	(14.36)
1+ Parent Prof.	-0.207**	-0.192**	-0.205**	-0.179**	-0.084**	-0.383**
	(155.34)	(139.31)	(133.72)	(99.19)	(70.47)	(103.65)
1+ Parent FB	0.403**	0.386**	0.380**	0.392**	0.465**	0.327**
	(337.30)	(319.42)	(265.87)	(228.26)	(162.71)	(243.07)
Father Educ	-0.002**	-0.002**	-0.002**	-0.002**	-0.005**	0.001**
	(20.93)	(19.46)	(17.37)	(15.18)	(23.33)	(7.43)
Mother Educ	-0.007**	-0.007**	-0.007**	-0.007**	-0.007**	-0.005**
	(57.73)	(56.39)	(48.31)	(46.02)	(41.87)	(33.94)
Same Ancestry	0.083**	0.076**	0.076**	0.077**	0.242**	0.053**
	(159.39)	(148.12)	(132.04)	(107.00)	(89.93)	(103.97)
Child Order	-0.004**	-0.004**	-0.002**	-0.005**	-0.017**	0.000
	(10.61)	(11.00)	(5.86)	(11.64)	(19.38)	(0.18)
Only Child	0.001*	0.002**	0.002**	0.002	0.002	0.006**
	(2.03)	(2.98)	(2.95)	(1.79)	(1.26)	(10.31)
Asian	0.168**	0.185**	0.183**	0.186**	(a)	0.210**
	(83.69)	(89.45)	(67.73)	(73.40)		(96.57)
Hispanic	0.276**	0.195**	0.194**	0.197**	(a)	(a)
	(222.62)	(153.17)	(110.62)	(117.81)		
Concentration	(a)	0.578**	0.557**	0.600**	0.361**	1.189**
		(131.23)	(96.21)	(105.86)	(74.33)	(65.23)
_cons	0.326**	0.312**	0.327**	0.306**	0.322**	0.442**
	(189.18)	(171.55)	(152.62)	(120.38)	(86.76)	(117.89)
R^2	0.57	0.58	0.58	0.58	0.50	0.36
N	**2,711,065**	**2,711,065**	**1,393,932**	**1,317,133**	**453,928**	**2,257,137**

Note: t-statistics in parentheses.
(a) variable not entered
* significant at 5% level, ** significant at 1% level
Source: American Community Survey, IPUMS 2005-2011

Appendix A: Definitions of Variables

Data Creation Notes: American Community Survey, 2005-2011

1. How parents are identified:

   An individual is identified as a parent (father or mother) if

   A. He/she is older than age 18 and a household head or spouse, and has a child in the household currently or even born.
   B. He/she is older than age 18, the relation variable designates him/her as a child of the household head, and the relation variable designated someone as a grandchild in the household.
   C. Note: It is unlikely, though not impossible, that there is a parent younger than age 18 with a child older than 5.

2. How grandparents are identified:

   An individual is identified as a grandparent (grandfather or grandmother) if

   A. He/she is age 35 or older, the household head or spouse, and the relation variable designates someone as a grandchild in the household
   B. He/she is age 35 or older, the relation variable designates them as a parent or step parent of the household head, and there are children in the household
   C. Note: It is unlikely, though not impossible, that there could be a grandparent younger than age 35 with a grandchild older than 5.

3. How race/ethnic origin variables are defined:

   a. Asian

A. A child is identified as Asian, and coded as 1, if he/she is designated as Asian by the binary "racasian" variable. However, the child is recoded and receives a 0 if "two or more races" or "other" is reported under the more detailed race question.

B. Additionally, a child receives a 1 if under ancestry, he/she has any of the following ancestries: Burmese, Cambodian, Chinese, Cantonese, Filipino, Indonesian, Japanese, Okinawan, Korean, Laotian, Hmong, Malaysian, Thai, Taiwanese, Vietnamese, Amerasian, Asian, or reports Asian Only for race combined with ancestry which is either American, United States, or Not Reported

b. Hispanic

   i. A child receives a 1 if under the question, "are you Hispanic", he/she has a value corresponding to an ethnicity other than "Not Hispanic"

c. Indigenous

   i. A child receives a 1 if under the race question, he/she has a value corresponding to "American Indian or Alaskan Native"

d. Black

   i. A child receives a 1 if under the race question, he/she has a value corresponding to "Black/Negro", i.e. African-American.

4. Age and Years Since Migration

   a. "Age" is defined as the child's age

   b. For foreign-born children, "Years since Migration" is calculated as the child's birth year minus year of immigration for those who immigrated

5. How *concentration* is defined:
    A. The linguistic concentration index is defined as the number of speakers of a language from a list of the twenty most common languages in a given metropolitan area, based on census definitions of metropolitan boundaries from the year 2000 (metarea). The determination of the number of speakers of given languages in a metropolitan area was done on the full sample, before restricting it to children. Further detail is provided in the ACS codebook and the IPUMS website.
    B. If a child was a monolingual English speaker, but there was a modal language spoken in the household different from English, the child was assigned a concentration value corresponding to this language.
    C. The 24 languages included in this calculation are: Spanish, Tagalog, Chinese, French, German, Vietnamese, Korean, Russian, Italian, Portuguese, Arabic, Hindi, Polish, Creole, Cantonese, Japanese, Mandarin, Greek, Hebrew, Navaho, Dutch, Pennsylvania Dutch, Yiddish, and Aleut-Eskimo.
6. Same Ancestry
    a. A child receives a value of 1 for "Same Ancestry" if the ancestry code of the father is equivalent to that of the mother.
    b. 6 codes relating to Mexican ancestry and 2 for Chinese ancestry have been combined.
    c. As a robustness check all codes for Latin American and Caribbean Spanish-speaking countries were combined.

7. Definition of New York City: New York City was defined to be composed of the following PUMAs (Public Use Microdata Areas):

    A. Bronx: 3701-3710

    B. Brooklyn: 4001-4018

    C. Manhattan: 3801-3810

    D. Queens: 4101-4114

    E. Staten Island: 3901-3903

    Source: Baruch College, City University of New York, Geoportal

Appendix B: Supplemental Tables – Not for inclusion in the published version

Table B1: Concentration distribution by Group: Percent of Children in Each Race/Ethnicity group in the Concentration Categories, excluding New York City

Concentration	Asian	Chinese	Korean	Cuban	Dominican	Mexican	Hispanic	Non-Hispanic
0%	31.25	27.51	30.83	30.76	11.12	24.14	28.24	89.58
<0.5%	18.50	21.74	25.83	0.37	0.35	0.15	0.40	3.56
0.5%-1%	15.33	19.12	10.41	0.18	0.16	0.05	0.14	1.45
1%-2%	13.90	11.52	22.39	0.98	1.07	0.41	0.60	1.46
2%-3%	7.24	7.71	5.98	2.53	3.13	0.84	1.30	0.84
3%-4%	4.59	3.08	3.73	1.32	1.32	1.00	1.12	0.50
4%-5%	3.53	5.45	0.20	3.11	4.18	1.61	1.93	0.43
5%-10%	5.07	3.12	0.18	17.80	33.86	6.40	9.30	0.98
10+%	0.58	0.74	0.45	42.95	44.81	65.40	56.96	1.20
Total	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**

Source: American Community Survey, IPUMS, 2005-2011

Table B2: Regression Analysis, excluding New York City
Whether a child speaks a language at home other than English

	Overall	Overall	Male	Female	Hispanic	Non-Hispanic
Female	0.010**	0.010**	(a)	(a)	0.020**	0.008**
	(26.80)	(26.45)			(16.64)	(21.55)
Foreign Born	0.057**	0.039**	0.036**	0.043**	0.003	0.061**
Grandparent	(25.65)	(17.44)	(13.91)	(14.23)	(1.00)	(16.59)
1+ Parent Prof.	-0.207**	-0.193**	-0.205**	-0.179**	-0.078**	-0.391**
	(148.10)	(135.61)	(127.07)	(99.84)	(59.97)	(101.03)
1+ Parent FB	0.402**	0.386**	0.380**	0.393**	0.500**	0.317**
	(317.38)	(299.30)	(251.82)	(220.29)	(178.41)	(229.65)
Father Educ	-0.003**	-0.003**	-0.003**	-0.003**	-0.005**	0.001**
	(26.34)	(23.07)	(20.19)	(16.68)	(24.17)	(7.06)
Mother Educ	-0.007**	-0.007**	-0.007**	-0.007**	-0.008**	-0.005**
	(62.85)	(60.77)	(49.03)	(47.47)	(41.80)	(33.68)
Same Ancestry	0.075**	0.069**	0.068**	0.069**	0.182**	0.050**
	(144.70)	(131.50)	(109.35)	(98.64)	(74.94)	(100.71)
Child Order	-0.004**	-0.005**	-0.003**	-0.006**	-0.017**	-0.001**
	(12.97)	(13.80)	(8.19)	(13.65)	(19.04)	(3.38)
Only Child	0.001	0.002**	0.002**	0.001	0.002	0.006**
	(1.09)	(2.96)	(3.26)	(1.52)	(0.93)	(10.93)
Asian	0.166**	0.180**	0.178**	0.182**	(a)	0.212**
	(76.75)	(82.08)	(61.86)	(67.45)		(93.67)
Hispanic	0.279**	0.202**	0.201**	0.203**	(a)	(a)
	(212.12)	(150.89)	(114.96)	(113.03)		
Concentration	(a)	0.583**	0.560**	0.608**	0.379**	1.343**
		(118.61)	(91.23)	(92.91)	(68.87)	(62.95)
_cons	0.335**	0.316**	0.332**	0.310**	0.360**	0.446**
	(178.98)	(161.79)	(143.02)	(117.0)	(91.56)	(112.24)
R^2	0.57	0.58	0.58	0.58	0.49	0.35
N	**2,595,812**	**2,595,812**	**1,334,757**	**1,261,055**	**426,041**	**2,169,771**

Note: t-statistics in parentheses
* significant at 5% level, ** significant at 1% level.
(a) variable not entered
Source: American Community Survey, IPUMS 2005-2011

Table B3: Regression by Ethnicity, excluding New York City
Whether a child speaks a language at home other than English

	Asian	Chinese	Korean	Cuban	Dominican	Mexican
Female	0.027**	0.025**	0.041**	0.022	0.020	0.018**
	(8.98)	(5.01)	(5.87)	(1.68)	(1.70)	(12.42)
Foreign Born	-0.008	0.077**	0.009	0.035	0.011	-0.001
Grandparent	(1.11)	(7.22)	(0.37)	(1.35)	(0.44)	(0.41)
1+ Parent Prof.	-0.232**	-0.174**	-0.164**	-0.187**	-0.078**	-0.065**
	(44.32)	(16.96)	(15.77)	(10.02)	(6.68)	(43.52)
1+ Parent FB	0.477**	0.434**	0.341**	0.241**	0.396**	0.525**
	(89.51)	(43.56)	(25.89)	(12.83)	(11.66)	(138.26)
Father Educ	0.005**	-0.000	0.006**	-0.012**	-0.008**	-0.005**
	(6.68)	(0.35)	(2.99)	(4.45)	(3.41)	(16.98)
Mother Educ	-0.013**	-0.006**	-0.012**	-0.004	-0.008**	-0.008**
	(20.02)	(5.61)	(5.04)	(1.40)	(2.64)	(34.02)
Same Ancestry	0.180**	0.289**	0.486**	0.282**	0.183**	0.190**
	(32.76)	(30.42)	(39.64)	(13.51)	(10.10)	(52.99)
Child Order	-0.016**	-0.026**	-0.032**	-0.001	-0.038**	-0.015**
	(6.49)	(6.27)	(4.58)	(0.06)	(5.41)	(15.42)
Only Child	0.035**	0.050**	0.035	0.043**	-0.023	-0.005**
	(7.78)	(7.95)	(3.98)	(2.89)	(1.58)	(2.63)
Concentration	0.204*	1.482**	1.517**	0.781**	0.805**	0.300**
	(2.18)	(8.93)	(2.63)	(16.35)	(7.24)	(47.90)
_cons	0.348**	0.255**	0.184**	0.518**	0.543**	0.331**
	(29.10)	(14.89)	(4.65)	(8.69)	(11.91)	(62.51)
R^2	0.20	0.36	0.45	0.37	0.29	0.50
N	**91,932**	**27,819**	**13,402**	**6,135**	**4,309**	**268,043**

Note: t-statistics in parentheses
*significant at 5% level, ** significant at 1% level
Source: American Community Survey, IPUMS 2005-2011

Table B4.1: Standardized Regression Coefficients
Whether a child speaks a language at home other than English

	Overall	Hispanic	Non-Hispanic	Native-Born	1.5 Generation
Female	0.012**	0.021**	0.013**	0.014**	0.007**
FB Grandparent	0.015**	0.004	0.023**	0.020**	-0.001**
1+ Parent Prof.	-0.097**	-0.065**	-0.129**	-0.090**	-0.089**
1+ Parent FB	0.420**	0.483**	0.403**	0.375**	0.440**
Father Educ	-0.019**	-0.047**	0.009**	-0.028**	0.004**
Mother Educ	-0.053**	-0.062**	-0.046**	-0.055**	-0.029**
Same Ancestry	0.088**	0.178**	0.088**	0.081**	0.150**
Child Order	-0.007**	-0.031**	0.000	-0.002**	-0.009**
Only Child	0.002**	0.002**	0.008**	0.001**	0.021**
Asian	0.085**	(a)	0.146**	0.085**	0.081**
Hispanic	0.193**	(a)	(a)	0.200**	0.223**
Concentration	0.151**	0.148**	0.130**	0.179**	0.046**
N	**2,711,065**	**453,928**	**2,257,137**	**2,565,890**	**145,175**

* significant at 5% level, ** significant at 1% level
Source: American Community Survey, IPUMS 2005-2011

Table B4.2: Standardized Regression Coefficients, by Ethnicity
Whether a child speaks a language at home other than English

	Asian	Chinese	Korean	Cuban	Dominican	Mexican
Female	0.026**	0.021**	0.042**	0.028**	0.027*	0.020**
FB Grandparent	-0.004**	0.047**	0.007**	0.007**	-0.004	-0.001
1+ Parent Prof.	0.160**	-0.112**	-0.113**	-0.063**	-0.073**	-0.065**
1+ Parent FB	0.297**	0.343**	0.263**	0.223**	0.270**	0.513**
Father Educ	0.037**	-0.013**	0.031**	-0.050**	-0.059**	-0.043**
Mother Educ	-0.112**	-0.053**	-0.054**	-0.013**	-0.063**	-0.069**
Same Ancestry	0.142**	0.266**	0.457**	0.247**	0.208**	0.169**
Child Order	-0.028**	-0.035**	-0.042**	-0.011**	0.065**	-0.029**
Only Child	0.031**	0.047**	0.031**	0.027**	-0.022**	-0.005**
Concentration	0.013**	0.072**	0.054**	0.327**	0.164**	0.123**
N	**98,452**	**31,437**	**14,113**	**10,669**	**6,942**	**274,815**

* significant at 5% level, ** significant at 1% level
Source: American Community Survey, IPUMS 2005-2011