Economic Assimilation and Skill Acquisition: Evidence from the Occupational Sorting of Childhood Immigrants*

Marigee Bacolod
US Naval Postgraduate School

Marcos A. Rangel
Duke University and BREAD

December 2015

Abstract

We study the economic assimilation of childhood immigrants to the United States. We show that the linguistic distance of one's mother tongue to English interacted with age at arrival is closely connected to occupational sorting in adulthood. We document that these result from different degrees of complementarity between English-learning potential at arrival and the acquisition of multiple skills demanded in the labor market (cognitive, physical, socio-emotional, and communication). Childhood immigrants that arrive at younger ages better assimilate, developing a bundle skills very similar to those supplied by observationally equivalent workers when adults. Childhood immigrants from English-distant countries who arrived after the primary school years develop advantages in tasks for which US-born workers and Anglophone immigrants are not well-suited. These patterns emerge even after we net-out the effects of formal education and are pertinent to men and women of all racial backgrounds. Consistent with the complementarity argument, linguistic distance and age at arrival also play a significant role on the choices of major among US-college educated immigrants.

Keywords: Immigration, assimilation, skills, English, occupation.

*Bacolod: mbacolod@nps.edu, Graduate School of Business and Public Policy, Naval Postgraduate School, 555 Dyer Road, Monterey, CA 93943, (831) 656-3302; Rangel (corresponding author): marcos.rangel@duke.edu; Sanford School of Public Policy, Duke University, 302 Towerview Rd., Rubenstein Building, Office 194, Duke Box 90312, Durham-NC, 27708, (919) 613-7340. We are grateful to Ingo Isphording and Sebastian Otten for sharing their data of the Levenshtein distance measure. Sections of the analysis were performed while Rangel was a visiting scholar at Princeton University's Program in Latin American Studies. He is thankful for their hospitality. We would also like to thank participants of the All-California Labor Conference and the Southern Economic Association meetings for helpful comments and suggestions. All errors are our own.

1 Introduction

Little is known about the link between immigrant assimilation and the ability to accumulate the different skills required in the U.S. labor market.[1] This is particularly surprising given the growing literature recognizing that workforce skills and abilities are multiple in nature and have sizeable direct effects on workers' wages (Heckman and Rubinstein, 2001; Knudsen et al., 2006; Heckman and Mosso, 2014; Heckman, Humphries and Kautz, 2014). We attempt to fill this knowledge gap in the present article.

In the context of immigration into the United States, we hypothesize that a child's linguistic development at arrival as one of the main drivers of differential investments not only in adulthood language abilities but also on cognitive, socio-emotional, and physical capabilities (above and beyond the acquisition of formal educational credentials). Our contribution highlights the complementary nature of observed human capital and skills' acquisition and the dynamic interplay between them – a concept that lies at the heart of influential work by Heckman and Cunha (2007, 2008). In simple terms, the initial endowment of a given skill has direct impact over the costs and benefits of investments in the acquisition of either the same or other skills. Hence, faced with such restrictions on the formation of multiple skills, some immigrants consciously elect to invest in brawn rather than in brain, in mathematics rather than in poetry, in science rather than in history, in logical reasoning rather than in persuasion, and so forth. Ultimately, it is this skills bundle that determines labor productivity, labor market success and economic assimilation of immigrants.

We bypass the central difficulty in assessing immigrants' multiple skills assimilation by assembling suitable data from the Dictionary of Occupational Titles (DOT) and Occupational Information Network (O*NET) merged with the U.S. Censuses 1990 and 2000, and the American Community Surveys (ACS) 2009-2013.[2] The DOT and O*NET offer detailed characterization of skill requirements for multiple occupations, and are based on ratings by trained occupational analysts regarding how jobs are performed in establishments across the nation. From these occupational skill requirements, we carefully compute four broad categories of worker skills: communication skills, cognitive skills, socio-emotional skills, and physical skills. We focus our empirical analysis by pooling these data for cohorts of childhood immigrants

[1] We acknowledge that "assimilation" and "integration" are more general processes the process by which immigrants become full-fledged members of their host societies (see Fukuyama, 2006; Portes and Rumbaut, 2006; Akerlof and Cranton, 2010). Here we focus on the human capital and worker-productivity elements of immigrants' economic assimilation only.

[2] See Ruggles et al. (2015).

who arrived in the United States between 1960 and 2005. We also utilize novel data on the Levenshtein distance of one's country of origin towards the English language to classify immigrants' Anglophone and non-Anglophone origins.[3] Finally, we capitalize on the insight of Lenneberg (1967) and other contributions to the child development literature (Pinker, 1994 and Birdsong, 2006) that point to mid-childhood as a major turning point in linguistic, cognitive and behavioral development to examine immigrants arriving in the US at different ages.

Examining these data with a difference-in-differences empirical strategy we find that children who arrive at an earlier age employ skills in adulthood that are similar to observationally equivalent workers born in the United States or from (close to) Anglophone origins. This is true when it comes to the communication skill requirements of occupations, but is also the case for cognitive, socio-emotional and physical skills dimensions. In contrast, those who arrive at relatively older ages (10 or more), and whose mother tongues are distant from English, are engaged in occupations that are markedly different from *both* those who migrated at younger ages and (close to) Anglophone individuals in general.[4] Childhood immigrants born in English-distant countries arriving after the primary school years are the ones who show absolute advantage in physical skills as adults (approximately 0.2 of one standard deviation on that skill distribution) and disadvantage not only in communication skills, but also on cognitive and socio-emotional ones (0.3, 0.2 and 0.2 of one standard deviation, respectively). Given the relative difficulty of learning a second language for child migrants from English-distant countries who immigrated after middle childhood, optimizing behavior leads these migrants to specialize in skills applicable to more physical occupations and simultaneously refrain from investing heavily on the accumulation of communication, cognitive and socio-emotional skills.[5] These effects are reduced to approximately half of their original size when we account for the influence of formal education attainment. We can assert nonetheless that initial language development and age-at-arrival continue to have significant interactive and independent impact on adult skill attainment – over and above the effects of schooling.

[3]This measure was originally developed by the Max Planck Institute of Evolutionary Anthropology in Germany to aid the study geographical linguistic diversity and has been previously employed by Isphording and Otten (2013) to examine international trade patterns, and by Adsera and Pytlikova (2012) to discuss international migration patterns.

[4]Throughout the article we consider that an individual's mother tongue is the main language spoken on her country of birth.

[5]Incidentally, the existence of these complementarities invites a reinterpretation of the use of linguistic endowment at arrival as an instrumental variable for adult English knowledge in the estimation of the latter's causal effects over either wages, employment and other socioeconomic outcomes (Bleakley and Chin, 2004, 2010; Guven and Islam, 2015) or educational attainment (Beck, Corak, and Tienda, 2012).

We analyze a subsample of college graduates to further examine the connection between optimizing behavior and specialization under linguistic endowment differences and skill accumulation among immigrants. For this group, we first show that the skill accumulation patterns seen for all immigrants are also present among these highly educated (and US trained) immigrants, particularly in the case of communication and socio-emotional skills. Second, by capitalizing on the information about college-major choices collected in the most recent waves of the ACS, we identify that college-educated childhood immigrants from English-distant countries who arrived in the United States after the primary school years are relatively more likely to major in Science, Technology, Engineering and Math (STEM) fields rather than in Social Sciences or Humanities when compared to (closer to) Anglophone immigrants. This is consistent with our reasoning that individuals facing different costs for the acquisition of additional skills end up directing effort towards specialization in systematic ways. This is likely to be true even before college, when teenagers choose vocational training or courses to take while in high-school or even when they decide how much effort to put in studying different disciplines, for example.[6]

It is important to note that our work also complements very recent contributions to the population studies' literature on occupational segregation and workplace concentration (Andersson et al., 2014; Stromgren et al., 2014; del Rio and Alonso-Villar, 2015). We believe it is useful to characterize the skill dimensions of jobs immigrants sort into, rather than occupation labels. Occupation labels are like a black box, are not sufficiently segregated nor lend themselves to a natural ordering, while skill dimensions are more informative particularly if the goal is to inform the policy-oriented debate on the types of human capital investments that should be enhanced or incentivized to aid the assimilation of immigrants. In sum, our study advances this broader literature on immigration by focusing on the skills assimilation of childhood immigrants by age-at-arrival and source country's linguistic distance to English in viewing the immigrant-native earnings differential. In documenting the skills assimilation of childhood immigrants, we demonstrate *how* the skills among those who immigrate as children may determine their economic success with respect to US-born workers and may ultimately influence the shape of the distribution of earnings in the population (immigrants and natives). In other words, we see our approach as focusing on the fundamentals of immigrants' economic assimilation.

The remainder of our paper is organized as follows. Section 2 reviews background issues and the related

[6] We see these as important avenues for future research. Rangel and Shi (2015) take the first step in that direction.

literature. Section 3 describes the data and measures used in the analyses. Section 4 presents our empirical specifications. Our analyses of the effects of linguistic distance and age at arrival over skill accumulation and education attainment of childhood immigrants is presented in Section 5. Section 6 concludes.

2 Background and related literature

Despite its importance, the process of child migrants' skill accumulation and assimilation into the U.S. labor market is still not well-understood.[7] While several studies have found socioeconomic progress across immigrant generations, there is also evidence of segmented assimilation and even downward mobility for recent immigrants.[8] Borjas (2013), for example, uses data from the 1970-2010 decennial censuses and finds a slowdown in earnings assimilation of more recent immigrants. The earnings disadvantage of immigrants (relative to US-born workers) who entered the U.S. before the 1980s narrowed by 15 percentage points in their first two decades after arrival. In contrast, those who arrived after the 1980s have a much lower rate of wage convergence.

Assimilation theory and other empirical studies suggest that acquisition of language fluency accounts for immigrants' faster wage growth compared to native-born workers. Indeed, Borjas (2013) traces the patterns of wage convergence to those of English fluency.[9] Cohorts that entered prior to the 1980s experienced a 12-percentage point increase in English fluency during their first decade in the United States, while the cohorts that arrived after the 1980s show only a 4-percentage point improvement. The estimation of returns to native language fluency among immigrants is intuitively appealing, but poses many empirical difficulties. First, language skills may be correlated with confounding factors such as unobserved innate ability, parental background or other accumulated (and unmeasured) skills. Second, typical measures for language skills are coarse and likely subject to measurement error. For example, in the U.S. Census the measure employed is a self-assessment of respondent's ability to speak English: very well, well, or not at all.

Previous studies address the problems of endogeneity and measurement error of language skills by employing instrumental variable strategies. In particular, Bleakley and Chin (2004) turn to the extensive literatures in cognitive sciences and psychobiology for the well-documented relationship between age and

[7] A notable exception, yet restricted to formal education acquisition, is Beck, Corak, and Tienda (2012).

[8] See, for example, Grogger and Trejo (2002), and studies reviewed in Borjas (1999).

[9] See also Angrist and Lavy (1997), Berman et al (2003), Chiswick and Miller (2002, 2007, 2010); Bleakley and Chin (2004, 2010), Dustmann and Fabri (2003) Dustmann and van Soest (2001,2002), and Guven and Islam (2015), among others.

language-acquisition facility. Their creative identifying variation for the wage return to language skills is an interaction of age at arrival with Anglophone vs. non-Anglophone country of origin. The implementation of this strategy essentially attempts to identify the language skills (labor productivity) from difference in the age-at-immigration gradient in English proficiency (wages) for immigrants born in English and non-English speaking countries.

However, the interaction of a child migrant's age at arrival and Anglophone origin has been shown to matter for adult attainment in schooling that has itself direct effects over wages (as recognized by Bleakley and Chin, 2004). For example, Beck, Corak and Tienda (2012) show that the probability of being a high-school dropout increases significantly each year for childhood immigrants who arrive after age eight. The multiple-dimension and dynamic human capital investment framework we base ourselves on argues against attributing differences in earnings between younger and older (at arrival) Brazilian immigrants relative to the difference between Jamaicans exclusively to language fluency in adulthood, however. According to such framework, language-acquisition facility is an endowment that operates as a catalyst for learning other skills. Therefore, when we examine adult outcomes such as wages, age at arrival and Anglophone origin (which indirectly represent initial endowment of language skills) should not be considered as independent of skills that are not included in the model. Doing so would bias estimates of returns to adult fluency since: (i) the acquisition of such skills is complementary in different degrees to language acquisition, and; (ii) these skills command themselves rewards in the labor market.

Compelling empirical evidence in economics, developmental psychology, and neurobiology make the case that different skills are malleable at different ages, and point to critical periods in child development (Knudsen et al, 2006). The child development literature itself proposes a major turning point during middle childhood (between ages 7 to 11) in cognitive skills development.[10] To provide a concrete example, middle-childhood is the period when children make the transition from "learning to read" (primarily decoding words) to "reading to learn" (for information and comprehension). Therefore, during this primary school-age period (or even before), a childhood immigrant's cognitive, language and socio-emotional skills are being crucially shaped, providing a base for later schooling, and eventually determining which skills individuals elect to develop before and during accrual of experience in the labor market.

[10] Jean Piaget, one of the first psychologists to study child cognitive development, calls this the "concrete operational stage," when children begin thinking logically about concrete events but still have difficulty with abstract concepts.

To interpret our findings, we invoke the conceptual model of skill formation developed in Cunha and Heckman (2007, 2008). They build a model that recognizes multiple stages of childhood, with inputs at different stages being complements (i.e., earlier investments affect the attractiveness of later ones). These features suggest that investments during early childhood yield relatively higher returns than in later childhood. Children who acquire greater stocks of (cognitive and socio-emotional) skills in early childhood are more efficient at learning in later childhood. In other words, skill begets skill. Consistent with the framework, we hypothesize that age at arrival and linguistic distance to English have an interactive effect on the skill formation of child migrants. Those who arrive prior to middle childhood have the highest rate of assimilation, and the differential will be greatest for those who come from linguistically close countries of origin. In particular, those with a lower level of initial language learning potential who arrive in late childhood or as adolescents will end up with relatively lower levels of communication skills as adults and, given the circumstances of the processes involved in multiple skill acquisition, will invest relatively more is skills whose acquisition is not complemented by language. The process of testing this hypothesis allows us to highlight a broader mechanism for immigrants' earnings assimilation that goes beyond language skills in adulthood. Incorporating these nuances is important. Our results provide evidence that immigrants "fine-tune" their optimization strategies with respect to investments in acquiring skills. The results have implications both for assimilation theory and immigration policy.

It is important to note that children in immigrant families – defined as people under age 18 who are foreign born or who live with at least one foreign-born parent – have doubled in the U.S. population since 1990 (Mather, 2009). Official estimates indicate a 90 percent increase in unaccompanied children crossing the US-Mexico border just between 2013 and 2014.[11] These trends clearly put the spotlight not only on immigration policy reforms but also on the U.S. public education system that has to absorb these child migrants.[12] In our view, if economic assimilation of migrants is to be expected, it will have to materialize itself in the form of child migrants' skills converging to levels observed among natives. The closing of skill gaps would then translate into more equitable adult labor market outcomes.

[11]Department of Homeland Security, "Statement by Secretary Johnson About the Situation Along the Southwest Border," (press release, September 8, 2014), www.dhs.gov/news/2014/09/08/statement-secretary-johnson-about-situation-along-southwest-border.

[12]Another important consideration is the increase in immigration flows from countries in Asia and Africa where the main language is not English. See Elo et al. (2015).

3 Data

Data for this study come from the 1990 and 2000 Censuses (5:100) and the 2009 to 2013 American Community Surveys (1:100) made available in the Integrated Public Use Microdata Series (IPUMS, Ruggles et al 2015), the Revised Fourth Edition of the Dictionary of Occupational Titles (DOT), and the Occupational Information Network (O*NET) 13.0 database. Occupation-specific skill measures from the DOT and O*NET are merged to individual-level records in order to characterize the skills of native- and foreign-born workers.[13] Finally, computed linguistic distances between English and the main language in the immigrant's country of birth are also merged to the main data set. Each of these sources is further described below.

3.1 Measure of Initial English Endowment

To capture English language facility at arrival in the U.S., we utilize the Levenshtein distance to measure the linguistic distance of one's country of origin (or mother tongue) towards the English language. This measure was developed and computed by the German Max Planck Institute of Evolutionary Anthropology to study geographical diversity of languages. It is based on a very simple way of measuring the similarity of character strings by counting the number of operations to transfer one character string into another (Levenshtein, 1966). Rather than directly contrasting words, the computations compare the phonic representation of pairs of words with the same meaning in two different languages.

The implementation is based on a specific phonetic alphabet which employs characters within the standard ASCII alphabet to represent common sounds. The contrast between words yields the number of sounds that have to be substituted, added, or removed to transfer the one word into the other (Holman et al. 2011).[14] It is the average dissimilarity within this set of words that is taken to be our measure for the linguistic distance between a given language and English.[15] Since lexical similarity may be influenced by chance (Bakker et al. 2009), such as an overlap in the phoneme inventories, this quantity is normalized by using the average disparity of all $N(N-1)/2$ pairings of words with different meanings.

[13] We restrict ourselves to occupations that are listed in all calendar years covered in our pooled data. In that way we do not address the impact of the emergence of new occupations over our outcomes of interest.

[14] Forty-one different symbols representing seven vowels and thirty-four consonants are employed. The words used in the approach are taken from the a 40-word list including words which are common in nearly all the world's languages, including parts of the human body or expressions for common things of the environment. See Swadesh (1952).

[15] As described by Isphording and Otten (2011), there is only one consonant that has to be substituted between the English word "*you*" and German word "*du*." Meanwhile, to transfer "*maunt3n*", which is the transcription of "*mountain*", into "*bErg*", one has to remove or substitute each consonant and vowel.

In essence, this measure of distance ($LDND$) gives an approximation of the number of cognates between English and another language. More cognates indicates languages are closer to having common ancestries. Therefore, a smaller Levenshtein distance indicates a higher probability that a language shares characteristics with English and is likely correlated with the ease of acquiring English as a second language.

Table A1, in the Appendix, shows the closest and furthest immigrant languages to English in our sample. Countries such as Canada and New Zealand have $LDND = 0$. The closest non-Anglophone immigrant language in our data is Vincentian Creole English ($LDND = 41.57$), while the furthest language is Vietnamese ($LDND = 104.06$). In an attempt to provide a better understanding of this scale we present in Figure 1 alternative illustrations of correlated quantities. First we estimate probit models using linguistic distance as the only predictor for Anglophone origin based on the binary classification of official languages according to *The World Almanac and Book of Facts* (1999) and adopted by Bleakley and Chin (2004). Then we plot the relationship between the linguistic distance measure and model-based projected probabilities. In the same figure we also plot the local polynomial relation between linguistic distances and scores of the 2005/2006 Test of English as a Foreign Language (TOEFL), which is applied all over the world.[16] The figure clearly reveals that our measure of linguistic distance does contain information that is relevant for our application. We proceed to use such measure to define multiple sub-groups in the analyses that follow.

3.2 Occupational Skills Databases

Occupational information in the Dictionary of Occupational Titles (DOT) and in the Occupational Information Network (O*NET) are the result of comprehensive studies of how jobs are performed in establishments across the nation and are collected from multiple sources: surveys filled by workers performing the job, members of trade and professional associations, and site visits by trained occupational analysts. We compute job skill measures as composites of such data, and merge them with worker-level information using a crosswalk matching DOT and O*NET occupation codes to 1990 Census occupation codes from the National Crosswalk Service Center included on all IPUMS data sets.

The period covered in our study overlaps well with occupational information from both the DOT and the O*NET. Occupation skills information in the Revised 4th edition of the DOT was collected between 1978 and 1990 and published in 1991. Meanwhile, the O*NET program replaced the DOT and began data

[16]We use the reports downloaded from the Educational Testing Service (ETS) web-page with tables that list average scores per country of test-taker's nativity.

collection in June 2001. The O*NET 13.0 database includes occupational skills collected between 2001 and 2007. While releases earlier than version 13.0 of O*NET exist, they mainly contain extrapolated data from the DOT.[17] Thus, in using both the DOT and the O*NET, we are able to capture the occupational tasks and skills of childhood immigrants who were in the labor force over 1990-2010.

Similar to previous studies that utilize information from occupational databases, it is not possible to simultaneously use all of the variables capturing job skills. High collinearity makes precise estimation impossible. We use the textual definitions of DOT and O*NET variables and the O*NET Content Model to construct interpretable measures of worker skills. These broad skill categories are: Communication or Language skills, Cognitive skills, Socio-Emotional skills, and Physical skills. These skill indices are created using principal component (factor) analysis. The indices are constructed from the first factor and are re-scaled to have a mean of 0 and a standard deviation of 1.[18] The specific DOT and O*NET variables included for each skill measure are described in Panels A and B of Table A2, respectively.

We turn to previous studies using the DOT (e.g. Bacolod and Blum, 2010; Bacolod Blum and Strange, 2009) for constructing indices from the DOT. For example, the measure of language skills from the DOT is constructed from variables such as *gedl*, with 5 levels to capture reading, writing, and speaking skills required to perform the job. At high *gedl* levels, workers are required to read literature and write critiques, while at low *gedl* levels, workers need only write and speak simple sentences. Similarly, several variables from O*NET Content Model are used to construct a language skill index. These relate to a worker's "developed capacities in language skills and verbal abilities" and come from responses to the survey question, "How important is ______ (e.g. the variable *Oral Expression*) to your current job?" Respondents rate the skill on a 1 to 5 scale, with 1 as "not important" and 5 as "extremely important." At the occupation (SOC) level, each O*NET skill is a weighted average of respondents' ratings (on average there are 31 raters per occupation).

Several variables from the O*NET are used to construct the Cognitive skills. These are the non-language related variables categorized under "Basic Skills" and "Complex Skills" in the O*NET Content Model, and relate to a worker's "developed capacities that facilitate learning or more rapid acquisition of knowledge

[17]Occupational analysts were asked to map occupational data from the DOT to the O*NET Content Model, a conceptual framework developed using ideas in organizational analysis. Approximately 100 occupations a year were gradually transitioned from extrapolated data. By version 13.0 of the O*NET database, occupational data collected between 2001 to 2007 from more than 128,000 workers in 95,000 establishments are included. (Source: U.S. Department of Labor 2008)

[18]The first factor accounts for 96 to 100% (O*NET) of variation in all the variables included in each index.

related to work performance.” A high value on the Cognitive skills index indicates higher cognition and that skills such as Critical Thinking, Mathematics, and Problem-Solving, are very important in carrying out the job. As with the Cognitive skills index from the O*NET, the DOT Cognitive skills index capture aspects of cognition.

Our measure of socio-emotional skills are constructed using variables that indicate the skills required for workers to perform their job in relation to people to achieve goals in the workplace. These variables indicate how aware a worker has to be of others’ reactions, adaptability, ability to work under stress, and persuasion, among others. Meanwhile, the Physical skills measures from the DOT and O*NET are constructed to indicate the physical demands required for job performance. For instance, in the DOT five levels are used to capture the degree of strength requirements as measured by the job’s involvement in standing, walking, sitting, lifting, and carrying objects.

3.3 Data on individual characteristics and immigration status

Our main sample is composed of workers aged 25 to 38 at the time of the interview for the 1990 and 2000 Censuses and the 2009 to 2013 American Community Surveys (ACS), and weights are adjusted to reflect differences in sampling between the former (5% sample) and the latter (1% sample). The main subgroup of interest is composed of childhood immigrants who were less than age 18 at the time of entry into the United States and for which we have well-defined country-of-birth information. We exclude immigrants from countries for which a linguistic distance could not be unequivocally determined, for example, “Americas, ns,” “USSR, ns,” and “Indochina, ns.” The sample includes prime-aged workers who came to the US as children and have been in the country for 8 to 38 years (more recent waves of childhood immigrants are excluded because they are still of school age, and most are not fully participating in the labor market). Importantly, in our sample of immigrants and native workers are similar with respect to locations of the experience-earnings profile.

We further define subgroups by exploring visual cues in the relationship between our measure of linguistic distance and performance on TOEFL exams. We identify three distinct types of countries: (i) Anglophone, for which linguistic distance is zero or smaller than 60 (local-averaged TOEFL above 100); (ii) Linguistically close, for which linguistic distance is between 60 and 80 (local-averaged TOEFL between 90 and 100), and; (iii) Linguistically far, for which linguistic distance is above 80 (TOEFL scores below

90). We also capitalize on the age at immigration to generate identifiers for immigrants that arrived in the United States early (before age 10) or late (after age 9). Finally we also single out a group of child immigrants that were born abroad (in Anglophone countries or not) to US-born parents, who are granted automatic citizenship but who have experiences abroad that may mimic those of other immigrants. In Table A3 we present descriptive statistics (English fluency, demographics, educational attainment and economic outcomes) for these groups. For reference in the first column we report mean and standard-deviation for US-born workers. The remaining columns present statistics for the subgroups listed above.

In terms of use of a second language at home and English fluency the patterns across columns are the ones we expected. Individuals from linguistically far and those immigrating at older ages are more likely to use a second language and underperform in terms of (self-reported) English ability. When we focus on racial and ethnic composition we see that Hispanics are more concentrated among those from linguistically far groups while Blacks are more frequently in the Anglophone ones. There is also a larger share of males among the former. In terms of educational attainment, earnings and poverty status what we observe (and confirm in econometric exercises below) is that both the linguistic distance and the age at immigration seem to have strong correlation with formal education attainment and measures of economic success.

Interestingly, in Table A4 we document that most of the patterns observed among all childhood immigrants are also present among those that were born abroad to US-born parents. This suggests that there is at least some impact of the environment surrounding these individuals before immigration to the United States. We further explore this reasoning below.

Tables A5 and A6 reproduce the analysis by focusing exclusively on the variables of main interest in the present article: skill indices (z-scores). Once again we observe strong gradients, with linguistic distance and age at immigration both associated with lower communication, cognitive and socio-emotional skills and greater physical skills' accumulation. Similar but more muted patterns characterize immigrants born abroad to US-born parents.

4 Empirical specification

The objective of our estimation strategy is to identify the role of English-learning potential in determining skill levels in adulthood of those who were childhood immigrants. Such evaluation corresponds to the estimation of differences in an individual's *skill* (outcome of interest) under two distinct conditions, high and low leaning potential. But for each childhood immigrant only one learning condition is observed. Therefore, we develop comparisons between them and groups that can plausibly illustrate what would have been the counter-factual trajectory of skill accumulation.

Our strategy involves difference-in-differences estimates and triple-differences estimates. The notation here is the one used in the causal-inference literature: let $skill_{ic}$ represent the realized outcome (e.g.: O*NET skill index) for immigrant i who was born in country c. Let AE_{ic} be an indicator function, assuming the unit value if the childhood immigrant enters the United States after age nine. In addition, let D_{ic}^{far} be the indicator of mother tongue's linguistic closeness to English.

Define $skill_{ic}(1)$ and $skill_{ic}(0)$ as the theoretical outcomes of individual i when exposed and non-exposed to lower English-learning potential, respectively. Consequently, the effect of the "treatment" (lower English learning potential) for individual i can be represented by the difference $skill_{ic}(1) - skill_{ic}(0)$. This leads us to theoretically represent the expected effect of increasing learning difficulties among older childhood immigrants from a linguistically far country by:

$$\tau^{far}(X) = E[skill_{ic}(1) - skill_{ic}(0) \mid X_{ic}, \theta_c, D_{ic}^{far} = 1, AE_{ic} = 1], \tag{1}$$

where X is a vector of **observed** individual demographic characteristics, and θ_c represents general characteristics of her country-of-birth.

The identification of such a parameter requires, therefore, a specific modeling of the (not directly observed) counter-factual function $skill_{ic}(0)$. While one alternative would be to contrast outcomes for the individuals coming from the same country before age 10, the relevance of age-effects on assimilation patterns might confound the inference. However, if we can safely assume that the assimilation's age-pattern of childhood immigrants is independent of the mother tongue (except for their effect over English-learning potential, of course), then a sample of comparable "non-treated" observations (which are equally affected by the alternative elements) can be used to net out effects of any confounding factors. Phrased in a slightly different way, the modeling of the counter-factual outcomes for the treated group can be based on the observation of a control-group (immigrants from close-to Anglophone countries). In this way, as long as both groups were to follow parallel paths if mother tongue had no age-specific effects over outcomes, the causal-inference can be rewritten in terms of a difference-in-differences parameter:

$$\begin{aligned}\tau^{far}(X) \quad = \quad & \left\{E\left[skill_{ic} \mid X_{ic}, \theta_c, D_{ic}^{far} = 1, AE_{ic} = 1\right] - E\left[skill_{ic} \mid X_{ic}, \theta_c, D_{ic}^{far} = 1, AE_{ic} = 0\right]\right\} \\ & - \left\{E\left[skill_{ic} \mid X_{ic}, \theta_c, D_{ic}^{far} = 0, AE_{ic} = 1\right] - E\left[skill_{ic} \mid X_{ic}, \theta_c, D_{ic}^{far} = 0, AE_{ic} = 0\right]\right\}.\end{aligned} \tag{2}$$

Therefore, let us consider a parametric specification of the counterfactual model. In particular, assume a linear version of the outcome for individual i from country c under a no-treatment status:

$$skill_{ic}(0) = \alpha_0 + AE_{ic}\alpha_1 + D_{ic}^{far}\alpha_2 + X_{ic} \cdot \beta + \theta_c + \eta_{ic}, \tag{3}$$

where the effects of immigrating after age nine (α_1), and observed covariates (β) are common to all individuals, and $\theta_c + \eta_{ic}$ collapse all unobservable characteristics (origin-level and individual-level). Considering constant treatment effects $\tau^{far}(X) \equiv \tau^{far}$, the following empirical model (in terms of the realized outcomes) can be explored:[19]

$$skill_{ic} = \alpha_0 + AE_{ic}\alpha_1 + D_{ic}^{far}\alpha_2 + X_{is} \cdot \beta + \left(D_{ic}^{far} AE_{ic}\right) \cdot \tau^{far} + \theta_c + \eta_{ic}. \tag{4}$$

From this parametric formulation, it is clear that the sufficient identifying assumption implicit in the choice of a control group can be represented by:

$$E[\eta_{ic} \mid \theta_c, X_{ic}, D_{ic}^{far}, AE_{ic}] = 0 \tag{5}$$

This condition would allow the consistent estimation of τ^{far} by standard ordinary least squares (including country-of-birth fixed-effects) on the pooled cross-sectional samples of the treatment and control populations, that is: immigrant populations from Anglophone, Linguistically close and Linguistically far countries arriving in the United States before and after age ten.

In order to confirm the main patterns in the data we also apply the same model to a group of immigrants who were born abroad to US-born parents and entered the country at different ages. Like a falsification

[19] The constant treatment effect is assumed here for expositional convenience. In the empirical section below we examine heterogeneity in treatment effects across genders and racial groups.

exercise the objective of focusing on these individuals is to examine if country of origin classifications are capturing more than we consider to be linguistic distance. The assumption is that those born to US-born parents would have exposure to English that makes linguistic distance associated with country of birth less relevant, but that they would still be subject to country-specific elements that affect skill accumulation (like the quality of schools or level and pattern of economic development, for example). The measure of treatment effects among these individuals can then be statistically compared to those from our main exercise, this is a triple-differences estimator, and assumes that effects of other covariates over is common across all sub-populations).[20]

5 Empirical findings

This section presents and discusses the evidence to support our main conclusions. First, we present regression estimates that capture the impacts over (self-reported) English fluency and educational attainment. In doing so we confirm earlier findings in the literature using our sample and motivate a more detailed look at human capital investments. We then present the analysis of different measures of skills using our regression estimates as well as providing regression-adjusted visual displays (based on local polynomial smoothing) of the patterns we uncover. We then conduct some basic sensitivity analysis and discuss findings based on an auxiliary sample of ACS respondents that are college graduates (holding a Bachelor's degree). Finally we present findings that examine the heterogeneity of the skill-acquisition patterns.

5.1 Educational attainment

In Table 1 we focus on the patterns formal education attainment using our data and econometric strategy. Columns [1] and [2] report the average levels for older and younger childhood immigrants belonging to six different groups defined by linguistic distance and ancestry, respectively. Panel A presents results regarding high-school graduation (or some post-secondary education) while Panel B lists the results regarding at least college-degree education. It is important to notice these correspond to conditional averages and refer to a non-Hispanic white male worker, of 38 years of age in the year 2000. The group that stands out is

[20] In our main empirical exercises we include observations US-born workers. Despite not directly contributing to the identification of the parameters of main interest, their inclusion still impacts the models' estimation of the effect of covariates like age, race and gender. Difference-in-differences and triple-differences estimates are not affected by their exclusion.

immigrants from linguistically far countries that immigrate after age nine. This observation is confirmed when estimating the age-at-arrival differences for each of the groups (columns [3] and [4]). Given the similarity between those from Anglophone countries and those from linguistically close ones we report difference-in-differences exercises considering the combination of the latter two as a reference group for those from linguistically far origins.

Our parameter of main interest is reported in column [5] suggesting that the impact of a lower endowment of English learning potential reduces formal education attainment by about 15 percentage points (-14.6) in the case of high-school graduation and 7 percentage points in terms of college degree (-6.7). These are substantial differences considering baseline values. It can also be seen that children born abroad (in the same set of countries) to US-born parents do not have dramatic reductions in schooling as a function of origin and age at immigration. In fact, the difference between these two amounts is statistically significant as reported in column [6]. Finally, in column [7] we show that findings are robust to the exclusion of US-born individuals from the regression estimation. The results presented here are in line with evidence presented by Beck, Corak, and Tienda (2012) using the 2000 Census data.

5.2 Workplace skills and assimilation

Table 2 presents the results that are central to our analyses. We have estimated difference-in-differences and triple-differences parameters by focusing on those from linguistically far origins arriving after age nine as a treatment group, and using both Anglophone and linguistically close immigrants as the control group. We present results based on both DOT and O*NET classification, and focusing on communication, cognitive, socio-emotional and physical skills. What we find is that there are substantial and significant negative impacts of lower English-learning potential at arrival over the acquisition of all the computed skill indices.

In columns [1] and [2] we show that those arriving after primary school years from countries for which the language is distant from English are employing less communication skills in their current jobs. These differences correspond to 0.30 (DOT) or 0.34 (O*NET) of one standard deviation lower skills than their counterparts (considering the population distribution of such trait). These differences are significant even if we consider a more stringent criterion which takes into account the large sample sizes utilized on the estimation (Schwarz's Bayesian criterion). Moreover, these differences are still sizeable even after we net out the potential country-specific age-at-immigration effects within sending countries, made possible by examining those immigrants that were born abroad to US-born parents (triple-differences). It is clear from these results that (older at entry) immigrants coming with lower linguistic endowments end up accumulating and employing relatively less communications skills.

The same pattern is present when we focus on non-language cognitive skills (columns [3] and [4]). For these mathematical and logical reasoning skills the effects range from -0.19 to -0.27 of one standard deviation depending on the classification used. As before, language-learning disadvantages at the time of immigration translate into relatively fewer skills (or reduced intensity of given cognitive skills). Columns [5] and [6] confirm that the same applies to socio-emotional skills, with negative effects corresponding to 0.23 (DOT) and 0.20 (O*NET) of one standard deviation in the population distribution of those traits.

Finally, columns [7] and [8] show the opposite pattern we focus on physical skills. Individuals arriving in the United States as children who have a lower endowment of English-learning potential end up accumulating relatively *more* physical skills than their counterparts from (close to) Anglophone countries. These relative differences correspond to 0.18 to 0.22 of one standard-deviation in the distribution of such skills as measured by DOT and O*NET classification systems, respectively. Taken together with the results above, we can safely say that these immigrants indeed developed a bundle of skills that is tilted in favor of physical rather than intellectual and socio-emotional skills. In rather simple terms, they seem to have invested in brawn rather than in brain.

In order to better illustrate the nuances of our findings, we examine graphically the relationship between the DOT/O*NET-based measures of skills and age-at-arrival for countries that are linguistically far and those that are close (or are Anglophone). This is presented in Figure 2 utilizing four panels. The solid lines display the mean adjusted skills' z-score of childhood immigrants from (close to) Anglophone countries. The dashed lines display the means for immigrants from linguistically far countries. DOT measures are represented by thicker lines. In order to derive these means, first we generate individual-level residuals from a regression of skills on age indicators, dummies for race and ethnicity, and interview years. The figures then plot the average residual calculated for each age-at-arrival and for the two types of origin-country, which are then combined using a local polynomial smoother.

The patterns that emerge are clearly compatible with the reasoning we offer. Occupational language skills are lower among immigrants from linguistically far countries than for those from (close to) English-speaking countries. Most importantly, it is clear that migrating at younger is associated with relatively better communication, cognitive and socio-emotional skills in adulthood, particularly for those whose mother tongue is very different linguistically from English. For the latter, an additional year's delay in arrival – starting at age 5, but particularly strongly by age 11– is associated with progressively less skills.

The bottom-right panel in Figure 2 focuses on the pattern for physical skills of childhood immigrants. Among immigrants from English-speaking countries, physical skills remain essentially flat with age-at-arrival. In striking contrast to the negative relationships between age-at-arrival vs. communication, socio-emotional, and cognitive skills, however, immigrants from non-English-speaking countries tend to progressively acquire relatively *more* physical skills the older they are when they arrive in the U.S.

Figure 3 summarizes these findings by portraying difference-in-difference patterns. This is obtained by contrasting individuals from (close to) Anglophone and linguistically far origins by first computing skill levels relative to those immigrants arriving before completing the first year of age. This within group difference is then combined with an across-group comparison per age-at-immigration level. What we plot is the result of these two operations. As before, all conditioning on demographic is performed before these are computed. We can see from the pattern emerging here that age of arrival clearly implies a pattern of skill accumulation across all four dimensions studied. While levels of these different skills are not directly comparable (units of measurement do not lend themselves to such comparisons) we can learn from slight differences in the age-at-arrival gradient. It is interesting to note, in particular, that age seem to have a stronger effect on physical and communication skills (and from an earlier life-stage, at least in the O*NET classification).

5.3 Sensitivity and sub-group analysis

It is important to emphasize that the skill assimilation profiles discussed above are similar to those observed for the accumulation of educational credentials. It is natural to ask, therefore, if these patterns are not reflecting the same finding. In order to provide evidence that the effects of age-at-arrival and linguistic distance on skills assimilation we measure go beyond effects that arise because of educational attainment we re-estimate our models either using educational attainment as controls or by focusing on subsamples defined by highest degree attained. This time we focus solely on difference-in-differences estimations based on all immigrants who are not from linguistically far origins as controls (including all those born abroad to

US-born parents).[21] In Table 3 (PANEL B) we present evidence that despite the fact that higher completed education levels *are* associated with greater cognitive, language, and social skills and fewer physical skills, education does not explain all the differences we observe across our "treatment" and "control" groups. The effects measured are approximately half the size of the original ones, but we can safely assert that initial language capital and age-at-arrival continue to have significant interactive and independent effects on adult skill attainment – over and above the effect of education.

This can also be seen in an alternative approach presented in PANELs C and D of Table 3, which is based on sample stratifications. We can see in PANEL C that in qualitative terms the impact of English-learning potential at arrival is still seen among those who are high-school graduates (but have no more education than that). Finally, and interestingly, we find in PANEL D that among college graduates the patterns of skill accumulation are still sensitive to linguistic endowment but also that most of the effects are concentrated on communication and socio-emotional skills. It is interesting to note that the disadvantages for these two skills is not seen for cognitive skills, suggesting that highly-educated immigrants may have invested quite selectively.

This latter finding has indeed motivated one additional empirical exercise focused on childhood immigrants with a US-college degree. In order to so we returned to the 2009 to 2013 ACS data and re-selected a sample based on all Bachelor degree holders who report their choice of major captured in that survey. We kept the restriction in terms of age at immigration but removed the upper limit on current age and working status in order to draw most of the data available.[22] We employ an econometric strategy identical to the one used for the measures of skill presented above in order to verify differential patterns of major choice among college educated immigrants.

Our findings are extremely interesting. In Table 4, we can show the statistical and substantive significance of the impact of English-learning potential at immigration with respect to the choice of STEM majors over Social Sciences and Humanities majors. Older immigrants from linguistically far countries are 5.4 percentage points more likely to graduate with a STEM major, and they are 4.6 percentage points less likely to undertake Social Sciences or Humanities training. These patterns are also depicted in Figure 4, revealing a remarkable similarity with the skill figures shown above. This is a novel result that indicates

[21]For reference we re-estimate the model with this new definition of control group for the whole sample on PANEL A and find results that are nearly identical to the ones discussed above.

[22]This new working sample has descriptive statistics presented in Table A7.

the role of "innate" English language ability for the choice of college major. We feel this has important implications, and that this reasoning can also be applied to the understanding course-choice in high-school or even effort allocation across disciplines. It also raises questions about the profound impact both English teaching for immigrants as well as dual language schooling can have in the long run.

We conclude our analysis showing in Table 5 that the patterns of skill accumulation seen in the overall population sample also hold across subsamples defined by gender and race. Once again we can see from the pattern emerging here that age of arrival clearly implies a pattern of skill accumulation across all four dimensions of skills, for men, for women, for Blacks/other and for White/Asians. The impact of English-learning potential at arrival seems relevant across all these well-defined (and important) demographic groups.

6 Summary and Conclusions

Our analysis of US Censuses/ACS, distance of one's mother tongue to the English language, and skill content of occupations (DOT and O*NET) reveal important fundamental elements of the economic assimilation of childhood immigrants in the U.S. labor market. In documenting assimilation of childhood immigrants at the skill level, we demonstrate *how* the skills among those who immigrate as children may determine their economic success with respect to US-born workers and may ultimately influence the shape of the distribution of earnings in the population. In this way our work is related to a large literature in Economics on the wage effects of immigration on the U.S. labor market which shows that the effect of immigration on the wages of native-born workers is little to non-significant (e.g. Borjas and Katz 2007, Card 1990, among many others) while being relatively large among previous immigrants (e.g. Cortes 2008, Ottaviano and Peri 2010). Their findings are compatible with the idea that immigrants and US-born workers – even those with similar education and work experience – possess unique skills that lead them to specialize in different occupations, and this pursuit of their advantages in skill leads, in equilibrium, to a mitigation of locals' wage losses from immigration.[23]

We show that children that arrive at an earlier age sort themselves as adults into tasks/occupations

[23] Recent studies have argued for this imperfect substitutability in the U.S. (Lewis 2012, Ottaviano and Peri 2010, Peri and Sparber 2009), Germany (D'Amuri, Ottaviano and Peri, 2010), Spain (Amuedo-Dorantes and De la Rican 2011), and the UK (Manacorda, Manning and Wadsworth, 2007).

that are more similar to the ones chosen by observationally equivalent to native workers. In contrast, those who arrive at older ages – particularly after the primary school years – from English-distant countries are in tasks/occupations that are more physically intensive relative to those who migrated at younger ages and to native-born workers. By the same token they are less likely to be employing cognitive, socio-emotional and communication skills when performing tasks in their jobs. The significance of this interaction of age-at-arrival and linguistic distance (which we call English-learning potential) in determining adult attainment of communication, cognitive, physical, and socio-emotional skills of child migrants also go above and beyond the effect of migrants' acquisition of formal education.

We view this as evidence for the complementary nature of observed human capital and skills' acquisition and the dynamic interplay between them. English-learning potential at arrival is a major determinant of differential investments in multiple skills. Specifically, child migrants' English facility at arrival complements investments in education and development of cognitive, socio-emotional, and language skills employed in the labor market, but without parallel enhancement of physical skills. Child migrants who arrive young, particularly prior to the conclusion of middle childhood, are faced with fewer restrictions in skill formation, and are thus more likely to assimilate to native outcomes. The opposite is true for childhood immigrants who arrive from English-distant countries after age 9.

This reasoning is confirmed by an investigation of college major choice, which suggests that lower endowments of English-learning potential at arrival directly influence the tendency to choose STEM concentrations over Social Sciences or Humanities (both of which require greater language and communication skills). That is to say, this is consistent with individuals facing different costs for the acquisition of additional skills directing effort towards specialization in very systematic ways.

Our work contributes to the recent population studies' literature on occupational segregation and workplace concentration among immigrants (Andersson et al., 2014; Strongen et al., 2014; del Rio and Alonso-Villar , 2015). We believe the empirical evidence presented here helps characterize the skill dimensions of jobs immigrants sort into, going much deeper than occupation labels. In that sense, our work is more informative to the policy-oriented debate on the types of human capital investments that should be enhanced or incentivized in order to aid the assimilation of immigrants.[24] It also complements other empirical studies on assimilation theory which suggest that acquisition of language fluency (and only that

[24] This includes school-level Limited English Proficiency programs, or dual-Language education, for example.

skill dimension) accounts for immigrants' faster earnings growth compared to native-born or Anglophone workers (Borjas, 2013; Bleakley and Chin, 2004; and Guven and Islam, 2015). We provide evidence that this is an incomplete view of the assimilation process. By looking at childhood immigrants' acquisition of human capital we provide evidence that these patterns have a lot to do with the fact that early language skill begets workplace skills.

References

[1] Adsera, Alicia, and Mariola Pytlikova (2012). "The Role of Language in Shaping International Migration." IZA Discussion Paper No. 6333.

[2] Akerlof, George A and Rachel E Kranton (2010). "Identity economics: How our identities shape our work, wages, and well-being." Princeton: Princeton University Press.

[3] Andersson, Fredrik and Monica Garcia-Perez, John Haltiwanger, Kristin McCue, and Seth Sanders (2014). "Workplace Concentration of Immigrants," Demography, 51(6): 2281-2306.

[4] Amuedo-Dorantes, Catarina and S De la Rica (2011). "Complements or Substitutes? Task Specialization by Gender and Nativity in Spain," Labor Economics, 18(5): 697-707.

[5] Angrist, Joshua and Victor Lavy (1997). "The Effect of a Change in Language of Instruction on the Returns to Schooling in Morocco," Journal of Labor Economics, 15(1): S48-S76.

[6] Bakker, Dik, Andre Muller, Viveka Velupillai, Soren Wichmann, Cecil H. Brown, Pamela Brown, Dmitry Egorov, Robert Mailhammer, Anthony Grant, and Eric W. Holman (2009). "Adding typology to lexicostatistics: A combined approach to language classification." Linguistic Typology, 13(1): 169-181.

[7] Beck, Audrey N, Miles Corak, and Marta Tienda (2012). "Age at Migration, Social Boundaries and the Adult Attainments of Child Migrants," Annals of the American Academy of Political and Social Science, 643: 134-159.

[8] Berman, Eli, Kevin Lang, and Erez Siniver (2003). "Language-Skill Complementarity: Returns to Immigrant Language Acquisition," Labour Economics, 10(3): 265-290.

[9] Birdsong, David (2006). "Age and Second Language Acquisition and Processing: a selective overview," Language Learning, 56(1): 9-49.

[10] Bleakley, Hoyt and Aimee Chin (2004). "Language Skills and Earnings: Evidence from Childhood Immigrants," The Review of Economics and Statistics, 86(2): 481-496.

[11] Bleakley, Hoyt and Aimee Chin (2010). "Age at arrival, English proficiency, and social assimilation among US Immigrants," American Economic Journal: Applied Economics, 2(1): 165-192.

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

[13] Borjas, George and Lawrence Katz (2007). "The Evolution of the Mexican-Born Workforce in the United States" in Mexican Immigration to the United States, ed George J Borjas, Chicago: University of Chicago Press, 13-56.

[14] Card, David (1990). "The Impact of the Mariel Boatlift on the Miami Labor Market," Industrial and Labor Relations Review, 43(2): 245-257.

[15] —- (2009). "Immigration and Inequality," The American Economic Review, 99(2): 1-21.

[16] Card, David and John DiNardo (2000). "Do Immigrant Inflows Lead to Native Outflows?" The American Economic Review, 90(2): 360-367.

[17] Card, David, and Ethan Lewis (2007). "The Diffusion of Mexican Immigrants during the 1990s: Explanations and Impacts," in Mexican Immigration to the United States, ed George J Borjas, Chicago: University of Chicago Press, 193-228.

[18] Chiswick, Barry R., and Paul W. Miller (2002). "Immigrant Earnings: Language Skills, Linguistic Concentrations and the Business Cycle." Journal of Population Economics 15(1): 31-57.

[19] Chiswick, Barry R., and Paul W. Miller (2007). "Computer Usage, Destination Language Proficiency and the Earnings of Natives and Immigrants." Review of the Economics of the Household 5 (2): 129-157.

[20] Chiswick, Barry R., and Paul W. Miller (2010)."Occupational Language Requirements and the Value of English in the US Labor Market." Journal of Population Economics 23(1): 353-372.

[21] Cortes, Patricia (2008) "The Effect of Low-Skilled Immigration on U.S. Prices: Evidence from CPI Data," Journal of Political Economy, 116(3): 381-422.

[22] Cunha, Flavio and James Heckman (2007). "The Technology of Skill Formation," American Economic Review, American Economic Association, 97(2):31-47.

[23] D'Amuri, Francesco, Gianmarco I.P. Ottaviano, and Giovanni Peri (2010). "The Labor Market Impact of Immigration in Western Germany in the 1990s" European Economic Review, Elsevier, vol. 54(4): 550-570.

[24] del Rio, Coral and Alonso-Villar, Olga (2015). "The Evolution of Occupational Segregation in the United States, 1940-2010: Gains and Losses of Gender-Race/Ethnicity Groups," Demography, 52(3): pp. 967-988.

[25] Dustmann, Christian, and Francesca Fabbri (2003). "Language Proficiency and Labour Market Performance of Immigrants in the UK." Economic Journal , Vol 113: 695-717.

[26] Dustmann, Christian, and Arthur van Soest (2001). "Language Fluency and Earnings: Estimation with Misclassified Language Indicators." The Review of Economics and Statistics 83 (4): 663-674.

[27] Dustmann, Christian and Arthur van Soest (2002). "Language and the Earnings of Immigrants," Industrial and Labor Relations Review 55(3): 473-492.

[28] Elo, Irma T., Elizabeth Frankenberg, Romeo Gansey, and Duncan Thomas (2015). "Language and the Earnings of Immigrants," Demography 52(5): 1513-1542.

[29] Fukuyama, Francis (2006). "Identity, Immigration, and Liberal Democracy," Journal of Democracy, 17(2): 5-20.

[30] Grogger, Jeff and Stephen Trejo (2002). "Falling Behind or Moving Up? The Intergenerational Progress of Mexican Americans," San Francisco, CA: Public Policy Institute of California

[31] Guven, Cahit and Asadul Islam (2015). "Age at Migration, Language Proficiency and Socio-economic Outcomes: Evidence from Australia," Demography, 52(2): pp. 513-542

[32] Haskins, Ron and Marta Tienda (2011). "Immigrant Youth," The Future of Children, Vol 21, Princeton: Princeton University.

[33] Heckman, James and Flavio Cunha (2007). "The Technology of Skill Formation," American Economic Review Papers and Proceedings, 97(2), 31-47.

[34] Heckman, James and Flavio Cunha (2008). "Formulating, Identifying, and Estimating the Technology for the Formation of Skills," Journal of Human Resources, 43(4), 738-782.

[35] Heckman, James, John Eric Humphries and Tim Kautz (2014). The Myth of Achievement Tests: The GED and the Role of Character in Amerian Life, University of Chicago Press.

[36] Heckman, James and Stefano Mosso (2014). "The Economics of Human Development and Social Mobility," Annual Review of Economics, 6(1), 689-733.

[37] Heckman, James and Yona Rubinstein (2001). "The Importance of Noncognitive Skills: Lessons from the GED Testing Program," American Economic Review Papers and Proceedings, 91(2): 145-149.

[38] Holman, Eric W., Cecil H. Brown, Soren Wichmann, Andre Muller, Viveka Velupillai, Harald Hammarstrom, Sebastian Sauppe, Hagen Jung, Dirk Bakker, Pamela Brown, Oleg Belyaev, Matthias Urban, Robert Mailhammer, Johann-Mattis List, and Dmitry Egorov (2011). "Automated dating of the world's language families based on lexical similarity." Current Anthropology, Vol. 52 (6): pp. 841-875.

[39] Isphording, I.E. and S. Otten (2013). "The costs of Babylon–Linguistic distance in applied economics," Review of International Economics, 21(2): 354-369.

[40] Juhn, Chinhui, Kevin Murphy, Brooks Pierce (1993). "Wage Inequality and the Rise in Returns to Skill," Journal of Political Economy, 101(3): 410-442.

[41] Knudsen, E., J. Heckman, J. Cameron and J. Shonkoff (2006). "Economic, Neurobiological, and Behavioral Perspectives on Building America?s Future Workforce." Proceedings of the National Academy of Sciences 103(27): 10155-10162.

[42] Lenneberg, Eric H. (1967). Biological Foundations of Language. Wiley.

[43] Lewis, Ethan (forthcoming). "Immigrant-Native Substitutability: the Role of Language Ability," in Immigration, Poverty and Socioeconomic Inequality, David Card and Steven Raphael eds.

[44] Manacorda, Marco, Alan Manning, and Jonathan Wadsworth (2007). "The Impact of Immigration on the Structure of Male Wages: Theory and Evidence from Britain."

[45] Mather, Mark (2009). Children in Immigrant Families Chart New Path. Population Reference Bureau. http://www.prb.org/pdf09/immigrantchildren.pdf Retrieved March 15, 2013.

[46] Ottaviano, Gianmarco I.P. and Giovanni Peri (2010). "Rethinking the Effect of Immigration on Wages," Mimeo UC Davis March 2010.

[47] Peri, Giovanni and Chad Sparber (2009). "Task Specialization, Immigration, and Wages," American Economic Journal: Applied Economics, 1(3): 135-159.

[48] Pinker, Steven (1994). The Language Instinct. New York, NY: William Morrow.

[49] Portes, Alejandro and Ruben G Rumbaut (2006). Immigrant America: A portrait. 3rd Ed. Berkeley and Los Angeles: University of California Press.

[50] Rangel, Marcos A. and Ying Shi (2015). "Skill begets Skills: Human Capital Investments and Specialization among Childhood Immigrants," Sanford School of Public Policy, unpublished working paper.

[51] Ruggles, Steven, J. Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek (2015). Integrated Public Use Microdata Series: Version 6.0 [Machine-readable database]. Minneapolis: University of Minnesota.

[52] Stromgren, Magnus, Tiit Tammaru, Alexander M. Danzer, Maarten van Ham, Szymon Marcinczak, Olof Stjernstrom, Urban Lindgren (2014). "Factors Shaping Workplace Segregation Between Natives and Immigrants," Demography, 51(2): 645-671.

[53] Schultz, T Paul (1998). "Immigrant Quality and Assimilation: A Review of the U.S. Literature," Journal of Population Economics, 11(2): 239-252.

[54] Swadesh, Morris (1952). "Lexico-Statistic Dating of Prehistoric Ethnic Contacts: With Special Reference to North American Indians and Eskimos." Proceedings of the American Philosophical Society, 96(4): 452-463.

[55] U.S. Department of Labor (2008). "O*NET Data Collection Program." *http://www.onetcenter.org/dlfiles/omb2008/SupportingStatement2.pdf*

[56] World Almanac and Book of Facts (1999). New York: World Almanac Books.

[57] Zhou, Min (2000). "Contemporary Immigration and the Dynamics of Race and Ethnicity" in America Becoming: Racial Trends and their Consequences, vol 1, Neil Smelser, William Julius Wilson, and Faith Mitchell, eds, 200-242.

Table 1: Educational Attainment by Immigrant Status, Linguistic Distance, and Age at Arrival

Pooled 1990/2000 Censuses and 2009-2013 ACS data

	Conditional mean						
	9<Age<18 at entry [1]	Age<10 at entry [2]	Diff. [3]	Diff. [4]	Diff-in-Diffs [5]	Triple Diffs [6]	Triple Diffs [7]
PANEL A: 1{High School Graduate or more}							
Non-US-born parents and…							
Ling. ``far'' origin	71.0	89.4	-18.38***	-15.22***	-14.61***	-13.14***	-12.85***
			(1.360)	(0.822)	(0.819)	(1.167)	(1.143)
Ling. ``close'' origin	95.0	95.6	-0.66	-0.83			
			(0.579)	(0.579)			
Angloph. origin	95.4	95.9	-0.42	-0.63**			
			(0.309)	(0.249)			
US-born parents and…							
Ling. ``far'' origin	90.2	96.0	-5.86***	-2.14***	-1.47		
			(0.983)	(0.749)	(0.959)		
Ling. ``close'' origin	96.2	95.8	0.35	0.17			
			(0.563)	(0.552)			
Angloph. origin	94.6	96.2	-1.64**	-1.46*			
			(0.801)	(0.766)			
PANEL B: 1{College Graduate or more}							
Non-US-born parents and…							
Ling. ``far'' origin	23.8	32.4	-8.61***	-7.24***	-6.70***	-8.47***	-9.04***
			(0.919)	(0.745)	(0.845)	(1.810)	(1.795)
Ling. ``close'' origin	38.6	33.3	5.31**	3.25			
			(2.072)	(2.010)			
Angloph. origin	38.9	41.4	-2.53***	-1.44*			
			(0.964)	(0.843)			
US-born parents and…							
Ling. ``far'' origin	40.0	40.9	-0.87	1.04	1.77		
			(1.247)	(1.203)	(1.700)		
Ling. ``close'' origin	34.0	34.7	-0.74	-1.90			
			(1.637)	(1.647)			
Angloph. origin	42.7	41.1	1.57	1.73			
			(1.944)	(1.878)			
Country-of-brith FE				Yes	Yes	Yes	Yes
Excludes US-born workers							Yes

*Notes: Sample is 4,830,465 observations (347,644 after excluding US-born workers). Standard-errors in parentheses under parameter estimates are clustered at the occupation-code level (similar results are obtained with two-way clustering at occupational code and country-of-birth following Cameron, Gelbach, and Miller, 2006). All models control for interview-year indicators; female, hispanic, asian, other race and black indicator variables. The influence of age at interview is accounted for semi-parametrically using indicator functions. Linguistic distance computation follows Max Planck Institute of Evolutionary Anthropology methodology discussed in the text and used by Isphording and Otten (2011). *** p<0.01, ** p<0.05, * p<0.1*

Table 2: Workplace Skills (z-scores) by Immigrant Status, Linguistic Distance, and Age at Arrival

Pooled 1990/2000 Censuses and 2009-2013 ACS

	Communication skills		Cognitive skills		Socio-Emotional skills		Physical skills	
	Diff-in-Diffs	Triple Diffs	Diff-in-Diffs	Triple Diffs	Diff-in-Diffs	Triple Diffs	Diff-in-Diffs	Triple Diffs
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
PANEL A: DOT classification								
Non-US-born parents and…								
Ling. ``far'' origin	-0.30***	-0.26***	-0.27***	-0.23***	-0.23***	-0.23***	0.20***	0.18***
	(0.031)	(0.043)	(0.034)	(0.045)	(0.030)	(0.041)	(0.032)	(0.041)
US-born parents and…								
Ling. ``far'' origin	-0.04		-0.04		-0.00		0.02	
	(0.031)		(0.032)		(0.028)		(0.032)	
PANEL B: O*NET classification								
Non-US-born parents and…								
Ling. ``far'' origin	-0.34***	-0.28***	-0.21***	-0.19***	-0.22***	-0.20***	0.26***	0.22***
	(0.035)	(0.040)	(0.033)	(0.043)	(0.029)	(0.038)	(0.027)	(0.043)
US-born parents and…								
Ling. ``far'' origin	-0.06*		-0.02		-0.01		0.04	
	(0.035)		(0.030)		(0.030)		(0.034)	
Country-of-brith FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

*Notes: Sample is 4,830,465 observations. Standard-errors in parentheses under parameter estimates are clustered at the occupation-code level (similar results are obtained with two-way clustering at occupational code and country-of-birth following Cameron, Gelbach, and Miller, 2006). All models control for interview-year indicators; female, hispanic, asian, other race and black indicator variables. The influence of age at interview is accounted for semi-parametrically using indicator functions. Linguistic distance computation follows Max Planck Institute of Evolutionary Anthropology methodology discussed in the text and used by Isphording and Otten (2011). *** p<0.01, ** p<0.05, * p<0.1*

Table 3: Difference-in-Differences Estimates for Workplace Skills (z-scores) after conditioning on Educational Attainment

Pooled 1990/2000 Censuses and 2009-2013 ACS

	DOT				O*NET			
	Communication skills	Cognitive skills	Socio-Emotional skills	Physical skills	Communication skills	Cognitive skills	Socio-Emotional skills	Physical skills
	Diff-in-Diffs	Diff-in-Diffs	Diff-in-Diffs	Diff-in-Diffs	Diff-in-Diffs	Diff-in-Diffs	Diff-in-Diffs	Diff-in-Diffs
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
PANEL A: Whole sample of immigrants without education attainment as controls (N=347,644)								
Non-US-born parents and…								
Ling. ``far'' origin	-0.29***	-0.25***	-0.23***	0.21***	-0.33***	-0.20***	-0.20***	0.26***
	(0.026)	(0.028)	(0.029)	(0.028)	(0.033)	(0.031)	(0.028)	(0.023)
PANEL B: Whole sample with education attainment as controls (N=347,644)								
Non-US-born parents and…								
Ling. ``far'' origin	-0.14***	-0.11***	-0.12***	0.11***	-0.18***	-0.08***	-0.11***	0.14***
	(0.017)	(0.018)	(0.019)	(0.019)	(0.020)	(0.020)	(0.018)	(0.016)
PANEL C: Sample of High-School graduates only (N=100,024)								
Non-US-born parents and…								
Ling. ``far'' origin	-0.19***	-0.16***	-0.13***	0.15***	-0.24***	-0.11***	-0.12***	0.19***
	(0.029)	(0.031)	(0.030)	(0.031)	(0.036)	(0.031)	(0.030)	(0.029)
PANEL D: Sample of College graduates only (N=98,696)								
Non-US-born parents and…								
Ling. ``far'' origin	-0.07**	-0.05	-0.13***	0.04**	-0.12***	-0.06	-0.12***	0.04*
	(0.030)	(0.030)	(0.034)	(0.015)	(0.024)	(0.041)	(0.028)	(0.021)
Country-of-brith FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

*Notes: Standard-errors in parentheses under parameter estimates are clustered at the occupation-code level (similar results are obtained with two-way clustering at occupational code and country-of-birth following Cameron, Gelbach, and Miller, 2006). All models control for interview-year indicators; female, hispanic, asian, other race and black indicator variables. The influence of age at interview is accounted for semi-parametrically using indicator functions. Education attainment controls are indicators for high-school, some college, and college or more. Linguistic distance computation follows Max Planck Institute of Evolutionary Anthropology methodology discussed in the text and used by Isphording and Otten (2011). *** p<0.01, ** p<0.05, * p<0.1*

Table 4: Difference-in-Differences Estimates for Choice of College Major (%)

Pooled 2009-2013 ACS

	STEM major	Social Sciences or Humanities major	Bio-Medical major	Business major
	Diff-in-Diffs	Diff-in-Diffs	Diff-in-Diffs	Diff-in-Diffs
	[1]	[2]	[3]	[4]
PANEL A: Whole sample of natives and childhood immigrants with a Bachelor's degree (N=2,714,985)				
Non-US-born parents and…				
Ling. ``far'' origin	5.39***	-4.59***	-0.89	0.65
	(1.301)	(1.052)	(0.772)	(1.032)
PANEL B: Sample of natives and childhood immigrants (under 16 at entry) with a Bachelor's degree (N=2,700,231)				
Non-US-born parents and…				
Ling. ``far'' origin	4.16***	-4.33***	0.04	0.14
	(1.222)	(1.122)	(0.735)	(1.015)
PANEL C: Whole sample excluding natives (N=122,508)				
Non-US-born parents and…				
Ling. ``far'' origin	4.92***	-4.69***	-0.81	1.04
	(1.184)	(1.076)	(0.784)	(1.007)
Country-of-brith FE	Yes	Yes	Yes	Yes
Mean among US-born:	*14*	*43*	*13*	*20*

*Notes: Standard-errors in parentheses under parameter estimates are clustered at the country-of-birth level. All models control for interview-year indicators; female, hispanic, asian, other race and black indicator variables. The influence of age at interview is accounted for semi-parametrically using indicator functions. Linguistic distance computation follows Max Planck Institute of Evolutionary Anthropology methodology discussed in the text and used by Isphording and Otten (2011). *** p<0.01, ** p<0.05, * p<0.1*

Table 5: Difference-in-Differences Estimates for Workplace Skills (z-scores), by demographic subgroups

Pooled 1990/2000 Censuses and 2009-2013 ACS

	DOT				O*NET			
	Communication skills	Cognitive skills	Socio-Emotional skills	Physical skills	Communication skills	Cognitive skills	Socio-Emotional skills	Physical skills
	Diff-in-Diffs	Diff-in-Diffs	Diff-in-Diffs	Diff-in-Diffs	Diff-in-Diffs	Diff-in-Diffs	Diff-in-Diffs	Diff-in-Diffs
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
PANEL A: Excludes natives, females only (N=155,990)								
Non-US-born parents and…								
Ling. ``far'' origin	-0.31***	-0.27***	-0.22***	0.13***	-0.39***	-0.24***	-0.24***	0.26***
	(0.039)	(0.037)	(0.046)	(0.032)	(0.060)	(0.053)	(0.049)	(0.039)
PANEL B: Excludes natives, males only (N=191,654)								
Non-US-born parents and…								
Ling. ``far'' origin	-0.26***	-0.23***	-0.23***	0.24***	-0.28***	-0.17***	-0.18***	0.24***
	(0.028)	(0.031)	(0.028)	(0.035)	(0.029)	(0.031)	(0.027)	(0.027)
PANEL C: Excludes natives, Blacks/Other race only (N=110,816)								
Non-US-born parents and…								
Ling. ``far'' origin	-0.29***	-0.26***	-0.20***	0.21***	-0.36***	-0.19***	-0.18***	0.28***
	(0.040)	(0.044)	(0.037)	(0.044)	(0.046)	(0.042)	(0.038)	(0.038)
PANEL D: Excludes natives, Whites/Asians only (N=291,439)								
Non-US-born parents and…								
Ling. ``far'' origin	-0.30***	-0.26***	-0.25***	0.23***	-0.34***	-0.21***	-0.22***	0.27***
	-0.026	(0.027)	(0.030)	(0.026)	(0.033)	(0.032)	(0.029)	(0.023)
Country-of-brith FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

*Notes: Standard-errors in parentheses under parameter estimates are clustered at the occupation-code level (similar results are obtained with two-way clustering at occupational code and country-of-birth following Cameron, Gelbach, and Miller, 2006). All models control for interview-year indicators; female, hispanic, asian, other race and black indicator variables. The influence of age at interview is accounted for semi-parametrically using indicator functions. Education attainment controls are indicators for high-school, some college, and colleg or more. Linguistic distance computation follows Max Planck Institute of Evolutionary Anthropology methodology discussed in the text and used by Isphording and Otten (2011). *** p<0.01, ** p<0.05, * p<0.1*

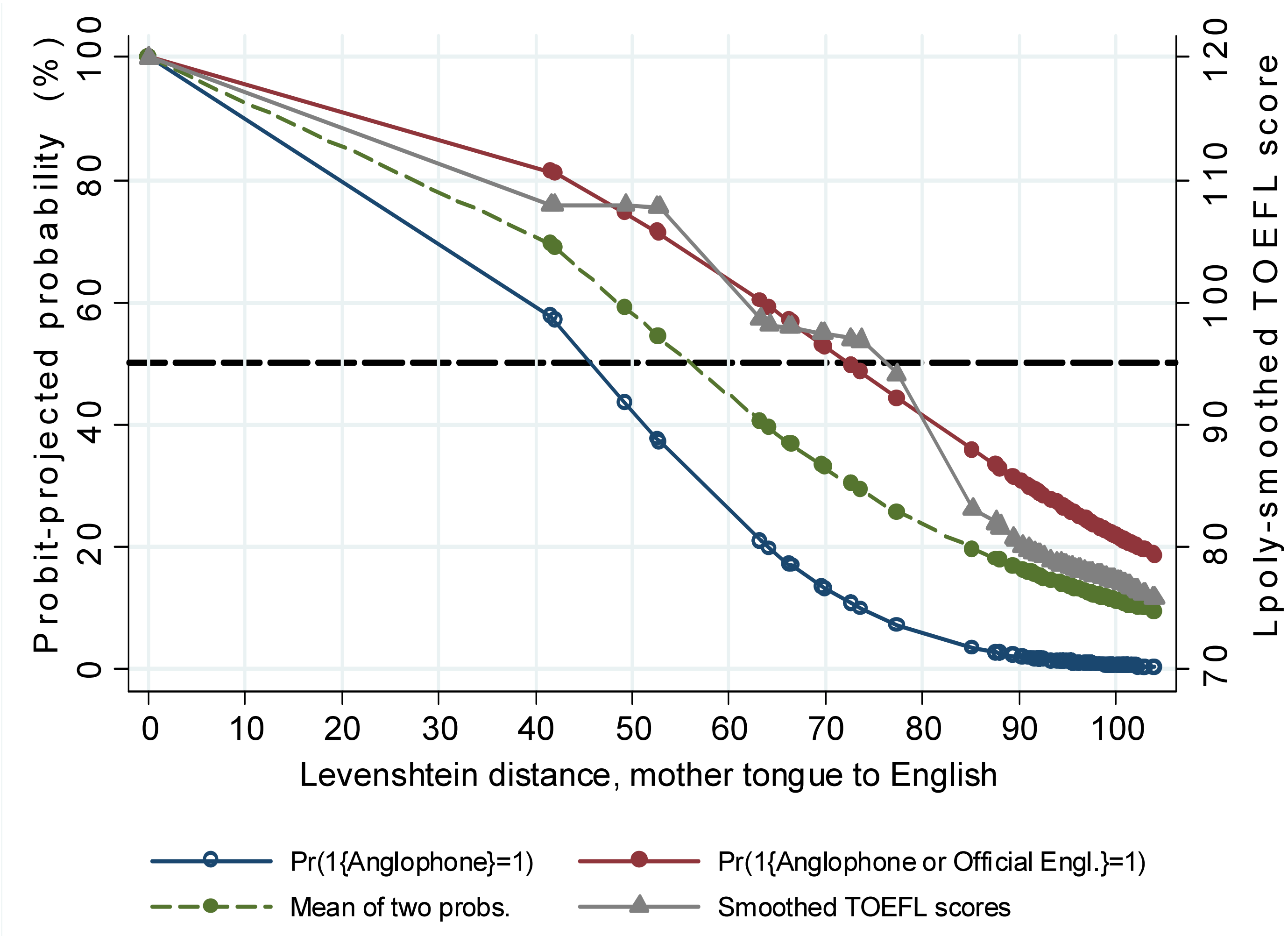

Figure 1: Projected relationship between Levenshtein Distance and TOEFL scores and Anglophone Classification (a la Bleakley and Chin, 2004)

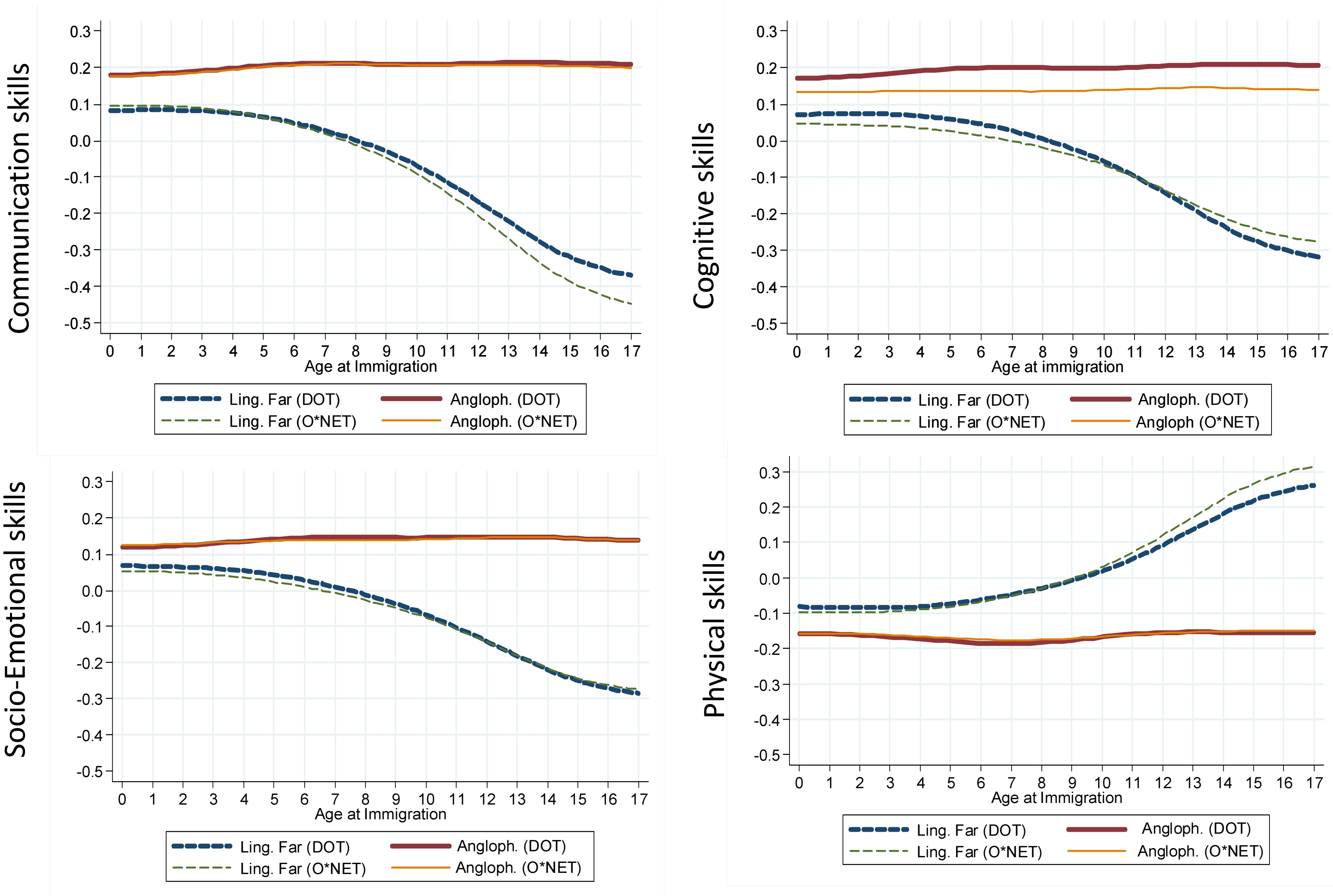

Figure 2: Relationship Between Skills and Age at Immigration, smoothed polynomials (regression-adjusted z-scores)

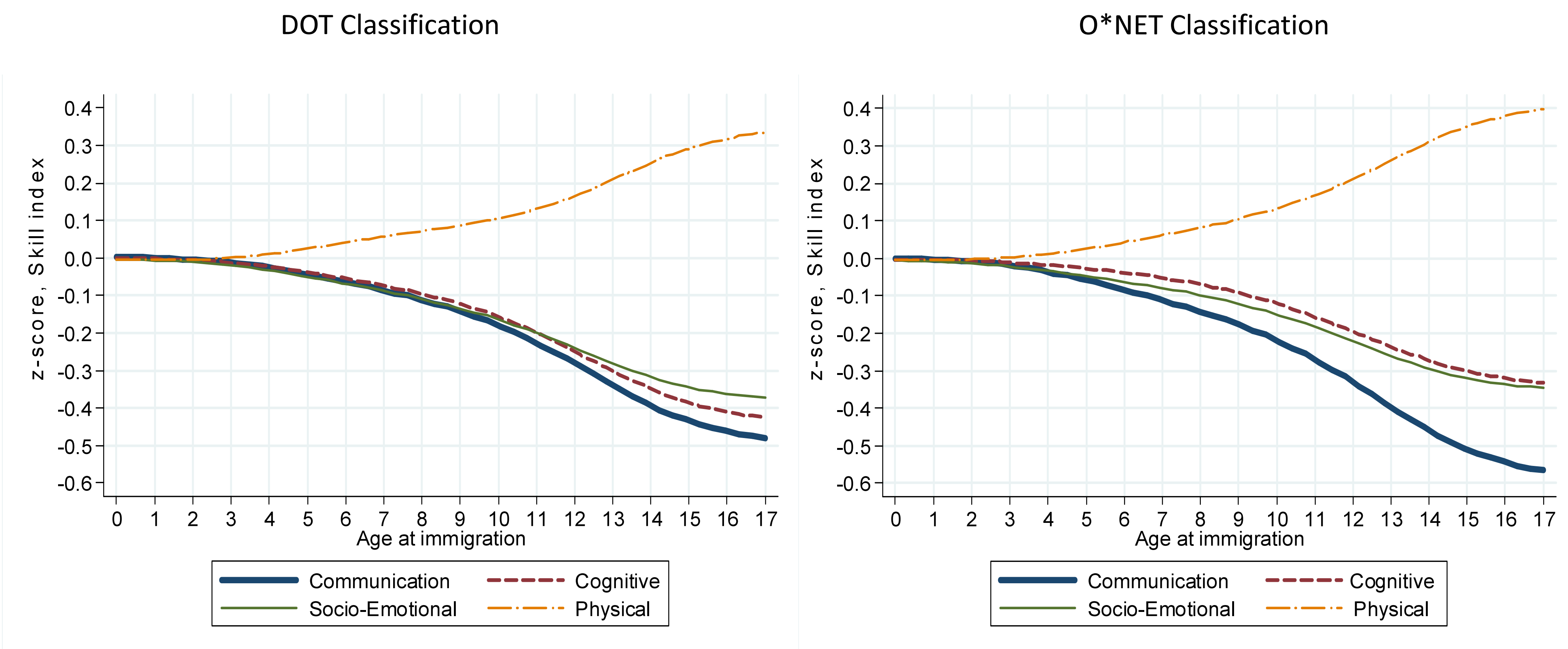

Figure 3: Difference in Differences - Workplace Skills and Age at Immigration, smoothed polynomials after regression-adjusted z-scores (age, gender, race and year of observation)

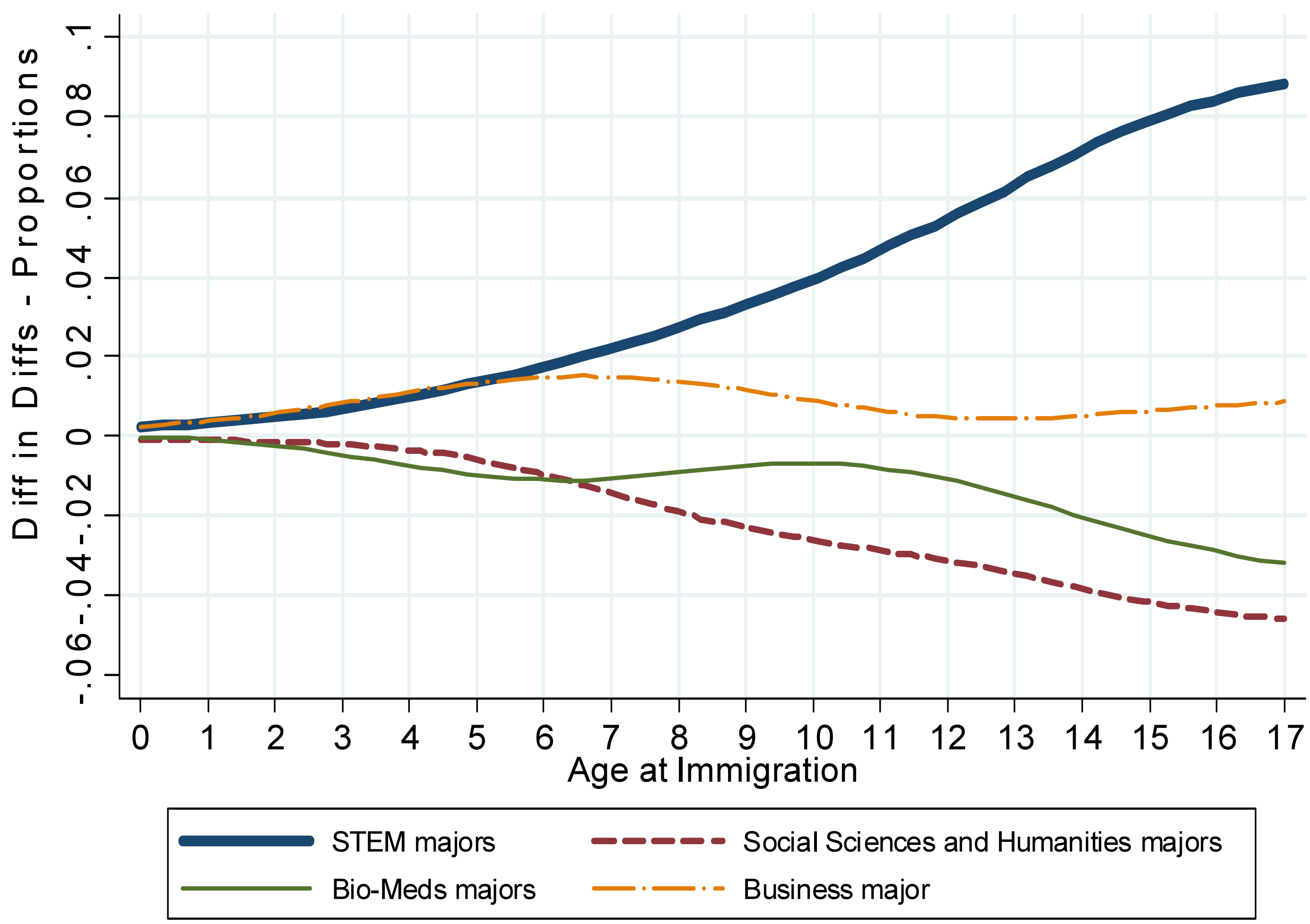

Figure 4: Difference in Differences - Choice of College Major and Age at Immigration, smoothed polynomials after regression-adjusted z-scores (age, gender, race and year of observation)

Table A1: Closest and Furthest Immigrant Levenshtein Distances to English

Closest		Furthest	
Country	LDND	Country	LDND
St Vincent	41.57	Vietnam	104.06
Jamaica	41.95	Somalia	103.03
Netherlands; Suriname	63.22	Burma/Myanmar	102.3
Aruba; Netherlands Antilles	63.22	Jordan	101.68

Table A2: Occupation-Based Measures of Immigrant Skills

Panel A. Variables from the DOT used for skill measures

DOT VARIABLES	DESCRIPTION
	Communication Skills Index:
gedl	general educational development in language required, from lowest to highest:
	(1) read at rate of 95-120 words per minute or vocabulary of 2,500 two- or three-syllable words; print and speak simple sentences
	(2) read at rate of 190-215 words per minute or passive vocabulary of 5,000-6,000 words; write compound and complex sentences, with cursive style, using adjectives and adverbs; speak clearly with correct pronunciation and proper tenses
	(3) read novels, magazines, safety rules and instruction manuals; write reports with proper format using all parts of speech; speak before an audience using correct English.
	(4) read novels, newspapers, encyclopedia; prepare business letters, reports; speak extemporaneously on various subjects, participate in discussions.
	(5) read literature, journals, reports; write editorials, manuals, critiques; conversant in persuasive speaking & debate.
aptv	segment of the population possessing verbal aptitude for the job: lowest 10% of popn; lowest third except bottom 10%; middle third; top 1/3 except top 10%; top 10% of popn.
talk	job requires talking and/or hearing
aptq	clerical perception (ability to proofread)
	Cognitive Skills Index:
data	complexity at which worker performs job in relation to data, from lowest to highest: comparing, copying, computing, compiling, analyzing, coordinating, synthesizing.
gedr	general educational development in reasoning required for job, ranging from being able to apply commonsense understanding to carry out simple instructions, to being able to apply logical or scientific thinking to wide range of intellectual and practical problems.
gedm	general educational development in mathematics required to perform job, from simple addition and subtraction, to knowledge of advanced calculus, modern algebra and statistics.
aptg	segment of the population possessing intelligence (or general learning ability) aptitude for the job: lowest 10% of popn; lowest third except bottom 10%; middle third; top 1/3 except top 10%; top 10% of popn.
aptn	segment of the population possessing numerical aptitude for the job: lowest 10% of popn; lowest third except bottom 10%; middle third; top 1/3 except top 10%; top 10% of popn.
	Physical Skills Index:
streng	degree of strength requirements of job as measured by involvement in standing, walking, sitting, lifting, carrying:
	(1) sedentary, sitting most of the time, walking and standing required only occasionally; exert up to 10 lbs of force occasionally to move objects
	(2) light work, more than sedentary; exert up to 20 lbs of force occasionally, up to 10 lbs frequently to move objects
	(3) medium physical work; exert 20-50 lbs of force occasionally, 10-25 lbs force frequently to move objects
	(4) heavy work; exert 50-100 lbs force occasionally, 25-50 lbs frequently, 10-20 lbs force constantly to move objects
	(5) very heavy physical demands; exert 100+ lbs of force occasionally, 50+ lbs frequently, 20+ lbs constantly.
climb	job requires climbing stairs, scaffolding, etc.
stoop	job requires stooping, kneeling, crouching
reach	job requires reaching, handling, fingering
hazard	environmental conditions on job
out	job involves outside activities
repcon	performing repetitive work
	Socio-Emotional Skills Index:
people	complexity at which worker performs job in relation to people, from lowest to highest: taking instructions; serving; speaking-signaling; persuading; diverting; supervising; instructing; negotiating; mentoring.
depl	dealing with people beyond instructions
dcp	responsibility for direction control planning
influ	influencing people in their opinions, attitudes
pus	performing under stress
sjc	adaptability to making decisions

Table A2: (con't)

Panel B. Variables from the O*NET used in constructing skill indices

O*NET VARIABLES	DESCRIPTION
	Communication Skills Index:
Reading Comprehension	Understanding written sentences and paragraphs in work related documents
Writing	Communicating effectively in writing as appropriate for the needs of the audience
Speaking	Talking to others to convey information effectively
Oral Comprehension	Ability to listen to and understand information and ideas presented through spoken words and sentences
Written Comprehension	Ability to read and understand information and ideas presented in writing
Oral Expression	Ability to communicate information and ideas in speaking so others will understand
Written Expression	Ability to communicate information and ideas in writing so others will understand
	Cognitive Skills Index:
Active Learning	Understanding the implications of information for current, future problem-solving & decision-making
Critical Thinking	Using logic & reasoning to identify the strengths, weaknesses of alternative solutions, conclusions or approaches to problems
Learning Strategies	Selecting & using training/instructional methods, procedures appropriate for the situation when learning or teaching new things
Mathematics	Using mathematics to solve problems
Monitoring	Monitoring/assessing performance of yourself, other individuals or organizations to make improvements or take corrective action
Science	Using scientific rules and methods to solve problems
Problem-Solving Skills	Identifying complex problems, reviewing related information to develop, evaluate options & implement solutions
	Physical Skills Index:
Dynamic Strength	Ability to exert muscle force repeatedly or continuously over time
Explosive Strength	Ability to use short bursts of muscle force to propel oneself as in jumping or sprinting or throw objects
Static Strength	Ability to exert maximum muscle force to lift, push, pull or carry objects
Trunk Strength	Ability to use abdominal & lower back muscles to support part of the body repeatedly or continuously over time without fatiguing
Stamina	Ability to exert physically over long periods of time without getting winded or out of breath
Dynamic Flexibility	Ability to quickly and repeatedly bend, stretch, twist, or reach with body, arms, and/or legs
Extent Flexibility	Ability to bend, stretch, twist, or reach with your body, arms, and/or legs
Gross Body Coordination	Ability to coordinate the movement of arms, legs, and torso together when the whole body is in motion
Gross Body Equilibrium	Ability to keep or regain body balance or stay upright when in an unstable position
	Socio-Emotional Skills Index:
Coordination	Adjusting actions in relation to others' actions
Instructing	Teaching others how to do something
Negotiation	Bringing others together and trying to reconcile differences
Persuasion	Persuading others to change their minds or behavior
Social Perceptiveness	Being aware of others' reactions and understanding why they react as they do

Table A3: Pooled 1990/2000 Censuses and 2009-2013 ACS data – Descriptive statistics

	US-Born	Born abroad to US parents	Anglophone age<10 at entry	9<age<18 at entry	Ling. Close age<10 at entry	9<age<18 at entry	Ling. Far age<10 at entry	9<age<18 at entry
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Sec. language	0.08	0.18	0.13	0.12	0.32	0.57	0.80	0.93
	0.27	0.38	0.34	0.32	0.47	0.50	0.40	0.25
English fluent	0.99	0.97	0.98	0.98	0.97	0.95	0.85	0.52
	0.11	0.17	0.13	0.12	0.18	0.22	0.36	0.50
Linguist Dist	.	71.31	7.50	13.92	71.25	70.74	94.89	94.37
	.	33.00	16.08	19.75	3.15	3.48	4.10	3.84
Projected TOEFL	.	92.27	117.85	116.02	97.27	97.35	78.61	78.75
	.	14.83	4.60	5.65	0.53	0.57	1.21	1.13
Age at immig	.	4.06	4.77	13.84	4.01	13.67	4.76	14.36
	.	4.45	2.73	2.32	2.61	2.43	2.84	2.25
Naturalized	.	.	0.63	0.52	0.75	0.49	0.63	0.44
	.	.	0.48	0.50	0.43	0.50	0.48	0.50
Age	31.54	31.24	31.43	31.31	32.34	31.57	30.83	30.91
	4.01	4.01	4.02	4.08	3.96	4.03	4.01	4.03
White	0.88	0.82	0.56	0.29	0.92	0.90	0.59	0.63
	0.33	0.38	0.50	0.45	0.27	0.30	0.49	0.48
Black	0.10	0.07	0.33	0.58	0.04	0.06	0.03	0.04
	0.30	0.26	0.47	0.49	0.20	0.23	0.18	0.20
Asian	0.01	0.05	0.06	0.06	0.03	0.03	0.26	0.20
	0.10	0.22	0.24	0.23	0.16	0.17	0.44	0.40
Other race	0.04	0.08	0.05	0.07	0.02	0.04	0.27	0.32
	0.19	0.28	0.22	0.26	0.15	0.19	0.44	0.47
Hispanic	0.08	0.12	0.01	0.02	0.02	0.03	0.57	0.66
	0.27	0.32	0.12	0.12	0.15	0.17	0.50	0.47
Male	0.51	0.52	0.49	0.45	0.51	0.51	0.53	0.61
	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.49
Schooling (yrs)	13.59	14.00	14.15	13.84	13.79	14.03	13.34	11.72
	2.11	2.15	2.06	2.15	2.00	2.15	2.64	3.54
HS grad+	0.94	0.96	0.97	0.96	0.96	0.96	0.88	0.68
	0.24	0.19	0.17	0.20	0.19	0.20	0.32	0.47
Some coll+	0.61	0.71	0.73	0.68	0.68	0.70	0.58	0.38
	0.49	0.46	0.45	0.47	0.47	0.46	0.49	0.49
Coll. grad+	0.32	0.40	0.41	0.36	0.34	0.40	0.32	0.20
	0.46	0.49	0.49	0.48	0.47	0.49	0.47	0.40
Ln hourly earn.	2.56	2.57	2.67	2.63	2.65	2.64	2.53	2.38
	0.70	0.71	0.71	0.69	0.69	0.76	0.74	0.74
< 130% pov. line	0.09	0.09	0.08	0.09	0.08	0.09	0.13	0.21
	0.29	0.28	0.27	0.29	0.27	0.29	0.33	0.41
< 185% pov. line	0.18	0.16	0.14	0.18	0.15	0.15	0.23	0.36
	0.38	0.37	0.35	0.38	0.35	0.36	0.42	0.48
Sample	4,482,821	43,662	13,817	12,712	4,781	1,517	106,927	164,228

Notes: Means (line 1) and standard-deviations (line 2) reported. IPUMS weights are employed on estimations. Sample consists of all US-born and immigrant workers with observed hourly earnings and between 25 and 38 years of age when interviewed for the 1990, 2000, 2009, 2010, 2011, 2012 and 2013 editions of the Census/American Community Survey (ACS). Childhood immigrants only (under age 18 at entry). Anglophone origin classification uses linguistic distance computation following Max Planck Institute of Evolutionary Anthropology methodology used by Isphording and Otten (2011).

Table A4: Pooled 1990/2000 Censuses and 2009-2013 ACS data – Descriptive statistics, Individuals born abroad to American parents who immigrated to US during childhood

	Angloph. age<10 at entry [1]	9<age<18 at entry [2]	Ling. Close age<10 at entry [3]	9<age<18 at entry [4]	Ling. Far age<10 at entry [5]	9<age<18 at entry [6]
Sec. language	0.05	0.09	0.08	0.25	0.23	0.60
	0.22	0.29	0.27	0.43	0.42	0.49
English fluent	0.99	0.99	0.99	0.98	0.97	0.83
	0.08	0.12	0.11	0.13	0.17	0.37
Linguist Dist	1.04	5.42	72.35	72.17	95.71	94.87
	6.52	14.08	1.52	1.88	3.97	3.81
Projected TOEFL	119.70	118.45	97.10	97.12	78.43	78.64
	1.87	4.03	0.30	0.33	1.18	1.13
Age at immig	2.82	13.55	2.28	13.38	2.79	13.70
	2.44	2.43	2.20	2.32	2.51	2.41
Age	31.53	30.72	31.32	31.06	31.18	30.85
	4.02	4.13	3.96	3.93	4.03	4.06
White	0.89	0.70	0.87	0.83	0.78	0.72
	0.31	0.46	0.34	0.37	0.42	0.45
Black	0.07	0.24	0.09	0.11	0.05	0.05
	0.26	0.43	0.29	0.31	0.22	0.23
Asian	0.01	0.02	0.01	0.01	0.09	0.11
	0.10	0.13	0.10	0.11	0.29	0.32
Other race	0.03	0.05	0.04	0.05	0.13	0.21
	0.17	0.21	0.19	0.23	0.33	0.40
Hispanic	0.02	0.02	0.04	0.05	0.18	0.40
	0.14	0.14	0.20	0.23	0.38	0.49
Male	0.53	0.52	0.51	0.51	0.52	0.55
	0.50	0.50	0.50	0.50	0.50	0.50
Schooling (yrs)	14.20	14.14	13.91	13.86	14.11	13.56
	2.06	2.23	1.99	1.94	2.14	2.82
HS grad+	0.97	0.96	0.97	0.97	0.96	0.89
	0.16	0.20	0.17	0.17	0.18	0.32
Some coll+	0.73	0.72	0.69	0.68	0.73	0.64
	0.44	0.45	0.46	0.47	0.44	0.48
Coll. grad+	0.43	0.44	0.36	0.35	0.42	0.38
	0.50	0.50	0.48	0.48	0.49	0.49
Ln hourly earn.	2.61	2.56	2.54	2.52	2.59	2.51
	0.72	0.72	0.71	0.68	0.71	0.75
< 130% pov. line	0.07	0.10	0.09	0.10	0.08	0.12
	0.26	0.30	0.28	0.29	0.27	0.32
< 185% pov. line	0.14	0.18	0.16	0.17	0.15	0.22
	0.35	0.39	0.37	0.38	0.36	0.41
Sample	6,314	994	15,657	1,120	16,106	3,471

Notes: Means (line 1) and standard-deviations (line 2) reported. IPUMS weights are employed on estimations. Sample consists of all childhood immigrants born abroad to US parents with observed hourly earnings and between 25 and 38 years of age when interviewed for the 1990, 2000, 2009, 2010, 2011, 2012 and 2013 editions of the Census/American Community Survey (ACS). Anglophone country-of-birth classification uses linguistic distance computation following Max Planck Institute of Evolutionary Anthropology methodology used by Isphording and Otten (2011).

Table A5: Pooled 1990-2013 Census/ACS data - DOT and O*NET Skill Measurements

	US-Born	Born abroad to US parents	Anglophone age<10 at entry	9<age<18 at entry	Ling. Close age<10 at entry	9<age<18 at entry	Ling. Far age<10 at entry	9<age<18 at entry
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
DOT Communication	0.017	0.175	0.236	0.151	0.142	0.232	0.021	-0.432
	0.994	0.971	0.952	0.936	0.947	0.943	1.018	1.034
DOT Cognitive	0.015	0.165	0.212	0.110	0.148	0.241	0.008	-0.379
	0.994	0.993	0.985	0.973	0.971	0.979	1.032	1.038
DOT Socio-Emotional	0.014	0.121	0.168	0.105	0.102	0.164	-0.003	-0.348
	1.001	0.977	0.964	0.941	0.969	0.940	0.984	0.943
DOT Physical	-0.009	-0.153	-0.207	-0.176	-0.134	-0.207	-0.082	0.294
	0.996	0.916	0.885	0.892	0.914	0.874	0.964	1.086
O*NET Communication	0.020	0.166	0.236	0.183	0.142	0.191	0.011	-0.500
	0.990	0.930	0.903	0.890	0.922	0.924	1.011	1.095
O*NET Cognitive	0.013	0.135	0.158	0.072	0.082	0.155	-0.013	-0.303
	0.996	1.005	1.009	0.991	0.994	1.056	1.040	1.019
O*NET Socio-Emotional	0.014	0.128	0.153	0.095	0.082	0.148	-0.030	-0.327
	1.000	0.976	0.977	0.961	0.971	0.984	0.997	0.945
O*NET Physical	-0.012	-0.137	-0.172	-0.116	-0.147	-0.208	-0.069	0.334
	0.997	0.968	0.976	0.991	0.969	0.929	0.995	1.018
Sample	4,482,821	43,662	13,817	12,712	4,781	1,517	106,927	164,228

Notes: Means (line 1) and standard-deviations (line 2) reported. IPUMS weights are employed on estimations. Sample consists of all US-born and immigrant workers with observed hourly earnings and between 25 and 38 years of age when interviewed for the 1990, 2000, 2009, 2010, 2011, 2012 and 2013 editions of the Census/American Community Survey (ACS). Childhood immigrants only (under age 18 at entry). Anglophone origin classification uses linguistic distance computation following Max Planck Institute of Evolutionary Anthropology methodology adopted by Isphording and Otten (2011).

Table A6: Pooled 1990-2013 Census/ACS data - DOT and O*NET Skill Measurements, Individuals born abroad to American parents who immigrated to US during childhood

	Angloph.		Ling. Close		Ling. Far	
	age$<$10 at entry	9$<$age$<$18 at entry	age$<$10 at entry	9$<$age$<$18 at entry	age$<$10 at entry	9$<$age$<$18 at entry
	[1]	[2]	[3]	[4]	[5]	[6]
DOT Communication	0.239	0.195	0.137	0.103	0.223	0.029
	0.970	0.998	0.964	0.954	0.963	1.018
DOT Cognitive	0.230	0.181	0.132	0.102	0.208	0.023
	0.994	1.019	0.981	0.977	0.989	1.035
DOT Socio-Emotional	0.166	0.141	0.099	0.055	0.149	0.023
	0.974	1.000	0.983	0.944	0.969	0.983
DOT Physical	-0.175	-0.170	-0.116	-0.129	-0.197	-0.079
	0.913	0.889	0.934	0.888	0.892	0.958
O*NET Communication	0.221	0.221	0.142	0.107	0.206	-0.002
	0.918	0.906	0.928	0.919	0.912	1.018
O*NET Cognitive	0.193	0.176	0.115	0.053	0.161	0.021
	1.007	0.996	0.989	0.998	1.008	1.042
O*NET Socio-Emotional	0.182	0.165	0.108	0.057	0.153	0.018
	0.970	0.959	0.974	0.968	0.974	0.998
O*NET Physical	-0.167	-0.129	-0.103	-0.100	-0.184	-0.037
	0.966	0.991	0.968	0.957	0.960	0.995
Sample	6,314	994	15,657	1,120	16,106	3,471

Notes: Means (line 1) and standard-deviations (line 2) reported. IPUMS weights are employed on estimations. Sample consists of all childhood immigrants born abroad to US parents with observed hourly earnings and between 25 and 38 years of age when interviewed for the 1990, 2000, 2009, 2010, 2011, 2012 and 2013 editions of the Census/American Community Survey (ACS). Anglophone country-of-birth classification uses linguistic distance computation following Max Planck Institute of Evolutionary Anthropology methodology used by Isphording and Otten (2011).

Table A7: Pooled 2009-2013 American Community Surveys – College graduates (B.A. holders) only, Descriptive statistics

	US-Born	Ling. Close age<10 at entry	9<age<18 at entry	Ling. Far age<10 at entry	9<age<18 at entry
Sec. language	0.06	0.16	0.25	0.68	0.84
	0.23	0.37	0.43	0.47	0.36
English fluent	0.99	0.99	0.98	0.93	0.84
	0.08	0.11	0.14	0.25	0.36
Linguist Dist	.	54.49	36.37	96.42	96.69
	.	39.94	38.45	4.37	4.37
Projected TOEFL	.	99.40	106.60	78.19	78.12
	.	16.60	15.66	1.31	1.33
Age at Immig	.	2.81	13.65	4.58	13.81
	.	2.67	2.38	2.87	2.33
Age	48.91	46.30	46.50	39.58	40.63
	15.51	13.52	15.17	11.41	11.90
White	0.89	0.78	0.55	0.43	0.40
	0.32	0.41	0.50	0.50	0.49
Black	0.07	0.10	0.31	0.03	0.06
	0.26	0.30	0.46	0.18	0.24
Asian	0.02	0.06	0.08	0.43	0.45
	0.13	0.23	0.26	0.49	0.50
Other race	0.02	0.06	0.06	0.10	0.09
	0.14	0.24	0.24	0.31	0.28
Hispanic	0.04	0.06	0.06	0.32	0.27
	0.19	0.23	0.24	0.47	0.44
Male	0.48	0.49	0.45	0.47	0.50
	0.50	0.50	0.50	0.50	0.50
STEM major	0.14	0.16	0.18	0.19	0.28
	0.35	0.37	0.38	0.39	0.45
Soc. Sci. major	0.09	0.11	0.11	0.11	0.08
	0.29	0.31	0.31	0.31	0.27
Hum major	0.35	0.31	0.27	0.23	0.18
	0.48	0.46	0.44	0.42	0.38
Bio-Meds major	0.13	0.13	0.14	0.15	0.15
	0.33	0.33	0.35	0.36	0.35
Other major	0.09	0.10	0.09	0.10	0.07
	0.28	0.30	0.29	0.30	0.26
Sample	2,592,477	32,004	10,153	41,369	38,982

Notes: Means (line 1) and standard-deviations (line 2) reported. IPUMS weights are employed on estimations. Sample consists of all US-born and immigrant workers with observed hourly earnings and between 25 and older when interviewed for the 2009, 2010, 2011, 2012 and 2013 editions of the Census/American Community Survey (ACS). Childhood immigrants only (under age 18 at entry). Anglophone origin classification uses linguistic distance computation following Max Planck Institute of Evolutionary Anthropology methodology used by Isphording and Otten (2011).