Early patterns of skill acquisition and immigrants' specialization in STEM careers

Marcos A. Rangel[a,1,2] and Ying Shi[b,1]

[a]Sanford School of Public Policy, Duke University, Durham, NC 27708; and [b]Center for Education Policy Analysis, Stanford University, Stanford, CA 94305

Edited by Mary C. Waters, Harvard University, Cambridge, MA, and approved November 26, 2018 (received for review July 12, 2018)

We provide empirical evidence of immigrants' specialization in skill acquisition well before entering the US labor market. Nationally representative datasets enable studying the academic trajectories of immigrant children, with a focus on high-school course-taking patterns and college major choice. Immigrant children accumulate skills in ways that reinforce comparative advantages in nonlanguage intensive skills such as mathematics and science, and this contributes to their growing numbers in science, technology, engineering, and math (STEM) careers. These results are compatible with well-established models of skill formation that emphasize dynamic complementarities of investments in learning.

STEM | immigration | skill acquisition | dynamic complementarity | comparative advantage

A large number of social science studies and policy debates focus on the economic trajectories of immigrants in the United States. Recent work recognizes that occupational choices play an important role, but the origins of specialization are not well understood (1–4). While the majority of studies focus on low-skilled workers, some highlight the participation of college-educated foreign-born workers in science, technology, engineering, and math (STEM) fields and their subsequent contributions to innovation and productivity gains (5–9). We show that patterns of STEM specialization in skill acquisition already emerge during secondary and postsecondary education by studying the experiences of immigrant children in the United States.

We argue that patterns of specialization among childhood immigrants are compatible with the idea that individuals acquire skills in ways that reinforce their early relative inclinations. This comparative advantage argument underpins classic economic texts including David Ricardo's study of countries' production specialization (10) and Andrew Roy's original writings on occupational choices (11, 12). Our reasoning is theoretically grounded in well-established models of skill formation (13, 14). Children with relative skill endowments tilted toward the English language will generate higher net returns to investment in reading/writing compared with peers. They will also be more inclined to accumulate skills that complement such a relative advantage. According to this theory of skill acquisition, certain immigrants will reinforce their comparative advantages in non–language-intensive skills such as mathematics and science while attending US-based educational institutions.

We infer the effect of an early relative advantage in nonlanguage skills on immigrants' academic trajectories by relying on two key empirical consistencies established in neurobiology, psychology, sociology, and economics. First, proximity of the mother tongue to English confers a distinct advantage in English language acquisition (15, 16). Second, age at immigration shapes future linguistic fluency (17–20): Younger individuals have greater facility in acquiring language skills, and those who arrive before a critical age can attain English proficiency in adulthood. So *ceteris paribus*, older Dominican immigrants face additional barriers to English proficiency relative to Jamaican immigrants, for whom English is the native language. Apart from language ability, young children exposed to a foreign school system before immigrating will also have a relative advantage in non-English school subjects in which they have been trained (e.g., numeracy).

We capture immigrants' comparative advantage upon arrival in the United States, using a combination of immigration age and linguistic distance to English. Our empirical approach relies on a difference-in-differences specification and the following intuition. A naive juxtaposition of immigrants from different countries of origin will confound the effects of relative language-related skill advantages with other factors such as parental expectations, cultural familiarity, support networks, role models, and country-specific selectivity inherent to the migration process (21). To address this source of selection by country of origin, we rely on immigrant children who arrive in the United States at different points of their development cycle but come from the same country. Cultural capital that influences a child's upbringing should be equivalent and therefore netted out in this process. Arguably the resulting difference in the age gradient between two groups of countries according to their proximity to English could be explained by differences in relative exposure to English and non-English subjects—which is the comparative advantage at immigration we proxy for.

Data

We trace the trajectories of immigrant children who entered the United States before age 16 y in two nationally representative datasets: the 2010–2016 waves of the American Community Survey (ACS) and the National Longitudinal Study of Adolescent Health (Add Health). We further verify the robustness of our results using the 2010, 2013, and 2015 waves of the National

Significance

Do net returns to early investments in skill acquisition explain patterns of occupational specialization among immigrants? We design an empirical analysis of nationally representative data to show that age at immigration and mother tongue influence course-taking patterns in high school and college major choice. Immigrant children with greater relative endowments in nonlanguage skills specialize in math and science course credits over English and are more likely to pursue science-, technology-, engineering-, and math-intensive subjects in college. This reveals the early influence of comparative advantage over immigrant career choices and economic assimilation in the United States.

Author contributions: M.A.R. and Y.S. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]M.A.R. and Y.S. contributed equally to this work.

[2]To whom correspondence should be addressed. Email: marcos.rangel@duke.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1812041116/-/DCSupplemental.

Published online December 31, 2018.

Survey of College Graduates (NSCG). Outcomes on degree attained and course-taking history come from self-reports (ACS, NSCG) and high-school transcripts (Add Health). *SI Appendix* describes these surveys and the construction of our analytical samples in greater detail.

Measurement

Linguistic distance between one's mother tongue and English is represented by Levenshtein scores. This measure captures the similarity of character strings between phonic representation of pairs of words with the same meaning in two different languages (22). More cognates between English and the other language indicate that they are more likely to have common ancestries. Previous literature has shown that a smaller Levenshtein distance is most likely correlated with the ease of acquiring English as a second language (23).

ACS data provide a large and representative sample to illustrate the usefulness of Levenshtein distances. Fig. 1*A* plots its distribution across countries represented within our sample of childhood immigrants. The distribution of linguistic distance is bimodal, with a cluster of immigrant-sending countries who are relatively fluent in English on the left and another concentration of immigrants registering high levels of Levenshtein scores relative to English on the right. Fig. 1*B* relates this measure to English fluency reported by immigrants who were in the United States for no more than 1 y. The farther away the predominant mother tongue is from English, the smaller the share of recent immigrants that report speaking English well. There exists a clear step function marked by steep drops in English fluency past certain Levenshtein distance thresholds.* We map the thresholds emerging from these analyses to a categorical variable on linguistic distance that distinguishes between linguistically close countries (in which either linguistic distance predicts that more than 50% of recent immigrants speak English well or where English is an official language) and countries considered linguistically far from English (Levenshtein measure of 83 or above).

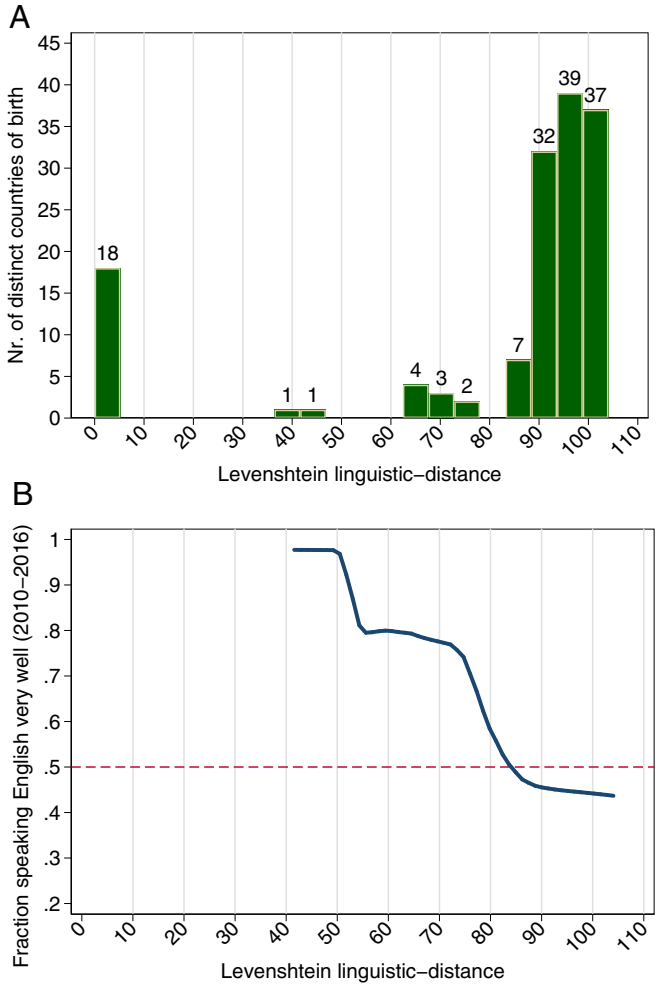

Fig. 1. (*A*) Distribution of Levenshtein measures of linguistic distance to English (unit of observation is country of origin within childhood immigrant sample). (*B*) Smoothed relation between linguistic distance and self-reported English fluency. Note that nonanglophone immigrants report within the first year of arrival in the United States. Immigrants from the 18 anglophone countries report an English fluency rate of 0.94 and are included under linguistic distance of 0. Data are from pooled 2010–2016 ACS and from ref. 23.

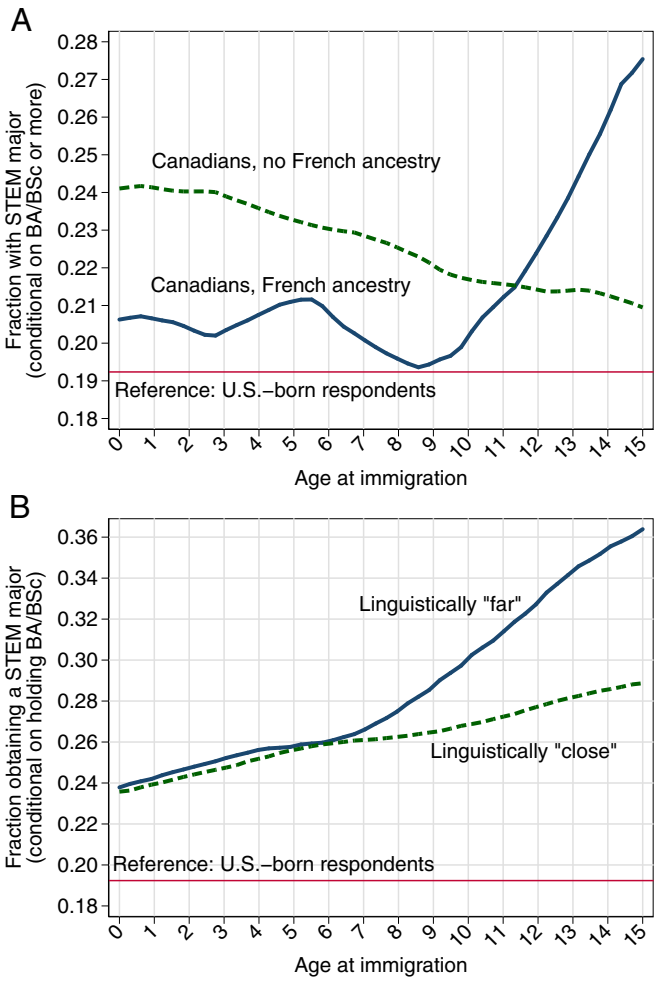

Fig. 2. (*A*) Smoothed relation between reported STEM college major (conditional on bachelor's degree) and age at immigration—Canadian immigrants only. (*B*) Smoothed relation between reported STEM college major (conditional on bachelor's degree) and age at immigration—linguistically far and linguistically close countries of birth. Note that data are from pooled 2010–2016 ACS. Childhood immigrants are 35 y or older at interview. ACS sample weights are used on all of the computations.

*In *SI Appendix*, Fig. S1 we further corroborate the existence of these steps employing data from the *Test of English as a Foreign Language* (TOEFL) together with our measure of linguistic distance. More detailed discussions and cross-validations of the Levenshtein measure in this context can be found in ref. 23.

In addition to linguistic distance, we rely on age at immigration to proxy for relative language skill disadvantages. Those who arrive before middle childhood have higher rates of language assimilation compared with later-arriving peers, particularly in the case of those that arrive from linguistically distant countries (15, 17). This absolute disadvantage in language skills has been shown by researchers to place immigrants at higher risk for academic underperformance and poorer physical and mental health while facing barriers to social assimilation and mobility more generally (17, 24–27). We begin our analyses by replicating select results on educational attainment and language fluency from the literature. Following previous studies, we use the interaction between age at arrival and linguistic distance to English to proxy for the costs of language acquisition. We segment the sample into two subgroups, linguistically close and far, using Levenshtein distance and the thresholds discussed above. *SI Appendix*, Fig. S2 suggests that childhood immigrants arriving at older ages are disadvantaged in terms of college attainment (*SI Appendix*, Fig. S2*A*) and English fluency in adulthood (*SI Appendix*, Fig. S2*B*). Most importantly, the age gradient appears distinctively steeper among those born in countries with a predominant language which happens to be linguistically distant from English.

Older immigrant children have also had more exposure to foreign educational systems in ways that can impact their skill accumulation in the United States. We depart from the previous literature in arguing that while the average immigrant has an absolute disadvantage along the dimensions described in *SI Appendix*, Fig. S2, he or she also holds a relative advantage in non–English-intensive subjects. According to this reasoning, those with a lower relative level of initial language learning potential should be expected to invest relatively more in skills whose acquisition is not complemented by English knowledge, conditional on the years of schooling attained. This would lead later-arriving students to gravitate away from the social sciences/humanities and specialize instead in mathematical and technical subjects when acquiring education. Following models of dynamic skill formation, we expect relatively small advantages early on to be reinforced through differential patterns of skill accumulation. Comparative advantage begets comparative advantage.

We make use of recently available college major choice information in the 2010–2016 ACS to illustrate these patterns. Fig. 2*A* begins with childhood immigrants from Canada and relies on the dual presence of francophone and anglophone Canadians. (We use an ancestry question in the ACS to identify French-Canadians within our sample.) Those with nonanglophone linguistic ancestry are more likely to show a strong gradient between age of arrival and postsecondary STEM orientation. (We follow definition of STEM major used by the US Department of Commerce and presented in ref. 28 which we reproduce in *SI Appendix*, Table S1.) Fig. 2*B* generalizes this reasoning across all countries in our sample and describes the smoothed relation between age of arrival and propensity to attain a degree in STEM, conditional on graduation. Fig. 2*B* indicates that differences start emerging as early as an age-of-arrival of 7 y. These patterns suggest that immigrants from a linguistically far country who were older at entry were much more likely to major in STEM—at rates that are twice the ones for the US-born population. (*SI Appendix*, Fig. S3 presents further illustrations of this pattern using different countries and regions.)

Methods

These descriptive exercises inform an empirical approach pioneered in the applied economics literature by Hoyt Bleakley and Aimee Chin (17) that relies on the interaction of immigration age with linguistic distance to proxy for English language potential and net out differential skill trajectories due to nonlanguage factors. We explore their reasoning in a difference-in-differences linear regression econometric specification detailed in *SI Appendix*. Importantly, this regression framework allows us to hold constant other characteristics that may be different for the individuals in our sample, including interview year, age at interview, race, and gender. We also net out all time-invariant country-of-origin specific effects using a fixed-effects framework.

Results

Table 1 reports coefficients from estimating our difference-in-differences models on the ACS sample. This approach subtracts disparities in skill investments between later- and earlier-arriving immigrants in linguistically close countries from disparities among immigrants from linguistically far origins. Assuming that factors influencing assimilation affect students of different ages similarly across countries of origin, the divergent outcomes for later-arriving immigrants from linguistically distant countries are more likely due to their relative nonlanguage skill advantages. We begin by examining the consequences of a relative advantage in nonlanguage skills on educational attainment. We find that immigrants entering the United States from a

Table 1. ACS: Effect of comparative advantage in nonlanguage material over college completion and major choice (percentage points)

Outcomes/immigrant groups	Under 10 y at immigration	10 y or older at immigration	Difference
i) College graduation, conditional on high-school graduation			
From linguistically far country	39.38 (5.68)	32.84 (5.90)	−6.54 (0.94)
From linguistically close country	43.32 (3.26)	40.00 (4.00)	−3.32 (1.55)
Difference in differences			−3.21 (1.81)
Difference-in-differences controls and country-of-birth effects			−0.90 (1.56)
ii) STEM major, conditional on college graduation			
From linguistically far country	25.51 (1.89)	33.89 (3.15)	8.38 (1.42)
From linguistically close country	24.65 (1.71)	27.98 (2.66)	3.34 (1.64)
Difference in differences			5.05 (2.16)
Difference-in-differences controls and country-of-birth effects			4.03 (1.24)
iii) STEM major, excludes far from anglophone countries with English as an official language			
Difference-in-differences controls and country-of-birth effects			4.55 (1.45)
iv) STEM major, does not consider official language classification for linguistic distance			
Difference-in-differences controls and country-of-birth effects			3.58 (1.43)

ACS childhood immigrant respondents are 35 y of age or older at interview. Sample is 286,869 for high-school graduates and 117,445 for college graduates (99,016 in section *iii*). All estimates are weighted and robust SEs are clustered at the country-of-birth level. Controls include age at interview indicators; year of interview indicators; male, Black, Hispanic indicators; and country-of-birth fixed effects.

linguistically far country who managed to complete high school are not significantly less likely to complete college than those arriving at younger ages. Notably, after controlling for college graduation our difference-in-differences estimates indicate that those coming with lower relative endowments of English skills are significantly more likely to major in STEM fields. The effect magnitude is 4.0 percentage points (SE: 1.2) or one-fourth of the average rate among US-born individuals. These results are robust to different classifications of countries with English as an official language (Table 1, sections *iii* and *iv*).

Fig. 3*A* takes advantage of detailed major classifications in the ACS to show results corresponding to STEM subgroups. These indicate that STEM increases are concentrated mostly in engineering, computer science, and applied mathematics, but are not significant in the physical and life sciences. As a counterpoint to the increase in STEM majors, we observe reductions in social sciences, business, and medical services degrees (Fig. 3*B*). These results suggest a specialization that shifts away from more language-intensive degrees toward science- and math-oriented

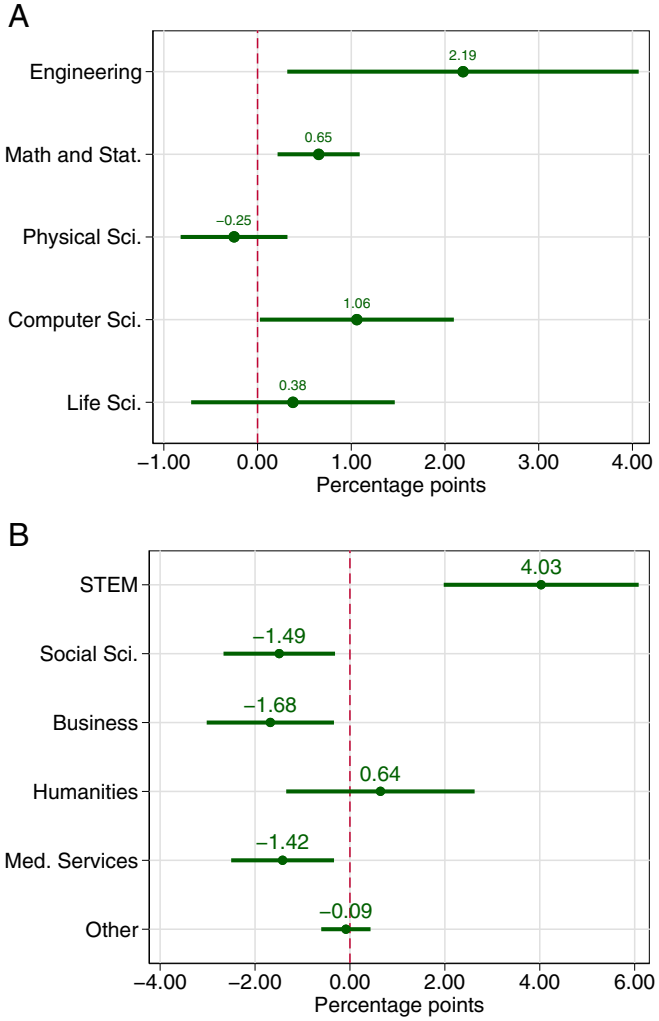

Fig. 3. (*A*) Difference-in-differences coefficients by groups of majors within STEM. (*B*) Difference-in-differences coefficients by fields. Note that ACS childhood immigrant respondents are 35 y of age or older at interview. Sample is 117,445 college graduates. All estimates are weighted and robust SEs are clustered at the country-of-birth level. Ninety-percent confidence intervals are depicted. Controls include age at interview indicators; year of interview indicators; male, Black, Hispanic indicators; and country-of-birth fixed effects.

subjects. To the extent that comparative advantage in English language during early childhood begets more investment in these capabilities later on, college enrollees are accumulating skills in a manner consistent with their original relative endowments.

The analyses on college major choice are informative, but they disguise the markers of differential skill investment during the precollege years. Major choice can be a symptom of earlier academic specialization, particularly in the case of STEM fields (29). To capture the acquisition of these inputs, we use nationally representative data from Add Health. The base sample comes from a subset of Wave I in-home interviews conducted during the initial period of 1994–1995, when respondents were in grades 7–12. An advantage of Add Health is the availability of high-school transcript data which document students' detailed academic trajectories. We focus on aggregate high-school credit accumulation over years 1–4 across multiple subjects such as math, English, science, and social science (including history). Since students may have completed different years of schooling, we ensure comparability by relying on the ratio of earned credits across two given subjects over years 1–4.

Table 2 shows the results of our main specification with outcomes as log-transformed ratios. They indicate that immigrants from anglophone countries arriving after age 10 y earn ∼21% more credits in math per credit in English, even after we net out the impact of age of arrival using immigrants from linguistically close countries. This relative credit intensity is similar in magnitude when considering both math and science credits. We find the same pattern in intensified math investment relative to social studies and history, which we consider more language intensive than other subjects, of 14%. Finally, we also see some evidence of relatively more intensive investment in physical education than math, which one can arguably say indicates increased investments in physical skills (22%) among high-school graduates who immigrated later from linguistically distant countries (*SI Appendix*, Table S2). This pattern would be consistent with a broader pattern of specialization and occupational choice, with some immigrant children investing relatively more in brawn (manually intensive jobs) than in brain (cognitively intensive jobs).

We also discuss the robustness of our results. First, our empirical approach so far posits that relative deficits in language-learning potential are the source of immigrants' comparative advantage in math- and science-intensive skills. Yet our conceptualization does not rule out the possibility that linguistically distant countries may simultaneously expose students to more or higher-quality STEM curricula at a young age. Therefore, we investigate whether differences in the quality of technical training fully explain the linguistic-distance age gradients we observe. We rely on data from the Program for International Student Assessment (PISA) examination of 15-y olds. To maximize the sample size of tested countries, we use math scores in the 2015 edition of the test under the reasonable assumption that cross-country differences in quality remain relatively stable across time. (We present the list of countries in PISA and ACS in *SI Appendix*, Table S3.) *SI Appendix*, Table S4 shows that including PISA math scores and their interaction with immigrant age in the regressions above generates an increase in the estimated difference-and-difference parameter to 5.59 (SE: 2.24) from 3.36 (SE: 2.18). This change reflects the fact that PISA math scores are inversely correlated with linguistic distance to English in our sample. The same pattern emerges when dichotomizing PISA scores into above or below the average US score. We find, therefore, that both linguistic distance and intensity of exposure to math in one's home country increase the chances that an older childhood immigrant would pursue a STEM major.

Second, we examine the possibility that our major choice estimates are explained by a pattern of selective immigration of

Table 2. Add Health: Effect of comparative advantage in nonlanguage material over course taking in high school

Outcomes/immigrant groups	Immigrant Under 10 y	Immigrant 10 y or older	Difference
i) Math vs. English			
From linguistically far country	−0.18 (0.023)	0.05 (0.046)	0.24 (0.061)
From linguistically close country	−0.07 (0.028)	−0.15 (0.056)	−0.08 (0.068)
Difference in differences			0.31 (0.090)
Difference-in- differences controls and country-of-birth FE			0.21 (0.093)
ii) Math and science vs. English			
From linguistically far country	0.48 (0.041)	0.68 (0.031)	0.21 (0.062)
From linguistically close country	0.58 (0.033)	0.51 (0.026)	−0.07 (0.048)
Difference in differences			0.28 (0.077)
Difference-in-differences controls and country-of-birth FE			0.20 (0.088)
iii) Math and science vs. social studies and history			
From linguistically far country	0.62 (0.040)	0.64 (0.053)	0.03 (0.054)
From linguistically close country	0.69 (0.067)	0.61 (0.084)	−0.07 (0.070)
Difference in differences			0.10 (0.087)
Difference-in-differences controls and country-of-birth FE			0.14 (0.066)

Note that Add Health respondents with nonmissing transcript data collected during the Adolescent Health and Academic Achievement study were born in a foreign country and immigrated before age 16 y. Covariates include gender, race, and indicators for age at interview. All estimates are weighted and robust SEs are clustered at the country-of-birth level. There are 829 observations in the comparison with English and 837 comparisons in the comparison with social sciences and history. FE, fixed effects.

high-ability Asians, who are overrepresented in our far-from-anglophone group. The argument is that US policy may have led to more Asian immigrants with educational and occupational qualifications that are predisposed toward STEM (selectivity), and their children may simply be following in the footsteps of their parents. (See discussion of the topic in ref. 21.) Our empirical method addresses most of these concerns by netting out country-specific effects. We formulate additional exercises to examine the results' robustness to the exclusion of Asian immigrants. *SI Appendix*, Table S5 reestimates our preferred specification using the full sample (*SI Appendix*, Table S5*A*) and multiple subsamples which exclude self-declared Asian respondents (*SI Appendix*, Table S5*B*), childhood immigrants from Asia (*SI Appendix*, Table S5*C*), and childhood immigrants from very-far linguistic distances (*SI Appendix*, Table S5*D*). The results of these exercises indicate that our main findings are still pertinent. Immigration from Asia appears to reinforce the patterns of specialization we observe, but evidence suggests that this source of selective immigration cannot fully account for our consistent results across multiple nationally representative samples. (In *SI Appendix* we present a more detailed investigation of other subsamples, in particular of immigrants from different cohorts.)

Finally, while our estimates are insulated from country-of-origin specific selectivity which affects immigrants both above and below 10 y of age, any differences in parental background or actions undertaken that correlate with the age of a childhood immigrant represent a threat to the interpretation of our results. Since our estimates are based on the outcomes of childhood immigrants during adulthood, little can be recovered regarding characteristics of parents and or investments they might have undertaken on behalf of their children. In an effort to cover some ground on this topic, we turn to data from the NSCG which captures parental education, reported visa status at immigration, and reported reason for immigration (descriptives in *SI Appendix*, Tables S6 and S7). In *SI Appendix*, Table S8, we find equivalent results for this smaller sample of immigrants: STEM majoring is more common for those with lower endowments of English skills at arrival, but these are not statistically significant with the exception of math and statistics. We then include additional controls in the model and examine the effects on difference-in-differences point estimates in *SI Appendix*, Table S9. We find that there are no noticeable changes in our point estimates for major choice despite differences among older immigrants from linguistically far countries in terms of visa type at entry (they are less likely to come with permanent status), motivation for migration (they are more likely to list educational plans), and parental education (parents less likely to have attended graduate or professional schools).

An investigation of potential differences in parental background using the Add Health sample further reinforces our original findings. We examine patterns of parental education, employment status, and occupation across immigration age and origin country (*SI Appendix*, Table S10). *SI Appendix*, Table S11 shows no statistically discernible differences across the immigration age gradient in terms of parents' educational attainment; employment status; and occupation in a professional, managerial, technical, or office setting at the time the they were interviewed during Wave I. Note that due to the limited sample size and relatively noisy estimates we cannot rule out more modest differences in household characteristics. When we reevaluate point estimates using these attributes as additional controls, findings in *SI Appendix*, Table S12 show that later immigrant arrivals from linguistically far countries still invest in significantly more math and science credits relative to language-intensive subjects.

Discussion

Immigrant assimilation is a longstanding topic of study among social scientists. (While each discipline focuses on different aspects of this process, they have by turns studied culture, language, discrimination, and barriers to skill acquisition. See canonical sociological perspectives in ref. 30 and a more modern and interdisciplinary approach in refs. 21, 31, and 32.) Our empirical approach builds on a broad literature in economics and sociology (17, 24, 33, 34) to examine immigrants' efforts to improve their social and material circumstances (35) via human capital accumulation. While this work is consistent with previous findings on the importance of language proficiency for educational trajectories and wage growth (17, 36, 37), our unique contribution is the focus on comparative advantage. We provide evidence that accounting for this dimension provides a

more complete view of the assimilation process in a context in which individuals sequentially make skill investment decisions (17, 38–45). Our findings corroborate the conceptualization of skill investments presented in refs. 13 and 14. Evidence across three nationally representative datasets shows that early patterns of specialization are a key factor contributing to certain immigrants' shift toward STEM subjects in high school and college. These comparative advantages lead to meaningful differences in skill accumulation before entering the labor market and shape the consequent contributions of immigrants to the educated labor force. (Interestingly, immigrants may also crowd out US-born individuals that would otherwise pursue STEM degrees. Evidence on this pattern is presented in ref. 46.)

Several labor market and policy implications follow. These patterns of skill acquisition shape immigrants' children's eventual career trajectories, and this in turn can affect their future economic wellbeing and the dynamics between immigrants' and natives' wage distributions. The role of comparative advantage cannot be overlooked when understanding why less-educated immigrants shift away from high language-intensity jobs (3) and more educated foreigners are well represented in STEM subjects (9). The focus on specialization can also inform education policies that seek to bridge skill gaps between immigrant and native children. Our findings illustrate the long-term impacts of immigration policies over US STEM-based innovations and show more broadly that education policies targeting STEM engagement could benefit by focusing on the early stages of a child's development.

ACKNOWLEDGMENTS. We are thankful to two anonymous referees for comments and questions which greatly improved the article and to Ingo Isphording and Sebastian Otten for sharing their data on the Levenshtein linguistic-distance measure. All errors are our own. We gratefully acknowledge funding from the Early Childhood Initiative (Duke University's Center of Child and Family Policy) and the Institute for Education Sciences (Award R305B130017).

1. Giovanni P, Sparber C (2009) Task specialization, immigration, and wages. *Am Econ J Appl Econ* 1:135–169.
2. Ottaviano GI, Peri G (2012) Rethinking the effect of immigration on wages. *J Eur Econ Assoc* 10:152–197.
3. Ottaviano GI, Peri G, Wright G (2013) Immigration, offshoring, and American jobs. *Am Econ Rev* 103:1925–1959.
4. Bacolod M, Rangel MA (2017) Economic assimilation and skill acquisition: Evidence from the occupational sorting of childhood immigrants. *Demography* 54: 571–602.
5. Hunt J, Gauthier-Loiselle M (2010) How much does immigration boost innovation? *Am Econ J Macroecon* 2:31–56.
6. Hunt J (2011) Which immigrants are most innovative and entrepreneurial? Distinctions by entry visa. *J Labor Econ* 29:417–457.
7. Maskus KE, Mobarak AM, Stuen ET (2013) Doctoral students and U.S. immigration policy. *Science* 342:562–563.
8. Peri G, Shih K, Sparber C (2015) STEM workers, H-1B visas, and productivity in US cities. *J Labor Econ* 33:S225–S255.
9. Hanson GH, Slaughter MJ (2016) High-skilled immigration and the rise of STEM occupations in U.S. employment (National Bureau of Economic Research, Cambridge, MA), NBER Working Paper 22623.
10. Ricardo D (1817) *On the Principles of Political Economy, and Taxation* (John Murray, London).
11. Roy AD (1951) Some thoughts on the distribution of earnings. *Oxf Econ Pap* 3: 135–146.
12. Heckman JJ, Honore BE (1990) The empirical content of the Roy model. *Econometrica* 58:1121–1149.
13. Cunha F, Heckman JJ (2007) The technology of skill formation. *Am Econ Rev* 97:31–47.
14. Cunha F, Heckman JJ (2008) Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *J Hum Resour* 43:738–782.
15. Isphording IE (2014) Disadvantages of linguistic origin—Evidence from immigrant literacy scores. *Econ Lett* 123:236–239.
16. Chiswick B, Miller P (2005) Linguistic distance: A quantitative measure of the distance between English and other languages. *J Multiling Multicult Dev* 26:1–11.
17. Bleakley H, Chin A (2004) Language skills and earnings: Evidence from childhood immigrants. *Rev Econ Stat* 86:481–496.
18. Long M (1990) Maturational constraints on language development. *Stud Second Lang Acquis* 12:251–285.
19. Kim K, Relkin N, Lee K-M, Hirsch J (1997) Distinct cortical areas associated with native and second languages. *Nature* 388:171–174.
20. Stevens G (2012) Age and second language acquisition among immigrants. *Handbook of US Immigration and Education*, ed Grigorenko EL (Springer, New York), pp 235–248.
21. Lee J, Zhou M (2015) *The Asian American Achievement Paradox* (Russell Sage Foundation, New York).
22. Levenshtein VI (1966) Binary codes capable of correcting deletions, insertions and reversals. *Sov Phys Dokl* 10:707–710.
23. Isphording IE, Otten S (2013) The costs of Babylon–Linguistic distance in applied economics. *Rev Int Econ* 21:354–369.
24. Tienda M, Haskins R (2011) Immigrant children: Introducing the issue. *Future Child* 21:3–18.
25. Clarke A, Isphording IE (2017) Language barriers and immigrant health production. *Health Econ* 26:765–778.
26. Bleakley H, Chin A (2010) Age at arrival, English proficiency, and social assimilation among US immigrants. *Am Econ J Appl Econ* 2:165–192.
27. Bleakley H, Chin A (2007) What holds back the second generation? The intergenerational transmission of language human capital among immigrants. *J Hum Resour* 43:267–298.
28. Langdon D, McKittrick G, Beede D, Kahn B, Doms M (2011) *STEM: Good Jobs for Now and for the Future* (US Department of Commerce, Washington, DC).
29. Altonji JG, Blom E, Meghir C (2012) Heterogeneity in human capital investments: High school curriculum, college major, and careers. *Annu Rev Econ* 4:185–223.
30. Gordon M (1964) The nature of assimilation. *Assimilation in American Life: The Role of Race, Religion and National Origins* (Oxford Univ Press, New York), pp 60–83.
31. Williams JA, Ortega ST (1990) Dimensions of ethnic assimilation: An empirical appraisal of Gordon's typology. *Soc Sci Q* 71:697–710.
32. Alba R, Nee V (1997) Rethinking assimilation theory for a new era of immigration. *Int Migr Rev* 31:826–874.
33. Hirschman C (2001) The educational enrollment of immigrant youth: A test of the segmented-assimilation hypothesis. *Demography* 38:317–336.
34. Glick JE, Bates L, Yabiku ST (2009) Mother's age at arrival in the United States and early cognitive development. *Early Child Res Q* 24:367–380.
35. Alba R, Nee V (2005) *Remaking the American Mainstream: Assimilation and Contemporary Immigration* (Harvard Univ Press, Cambridge, MA).
36. Guven C, Islam A (2015) Age at migration, language proficiency, and socioeconomic outcomes: Evidence from Australia. *Demography* 52:513–542.
37. Borjas GJ (2015) The slowdown in the economic assimilation of immigrants: Aging and cohort effects revisited again. *J Hum Cap* 9:483–517.
38. Chiswick BR (1978) The effect of Americanization on the earnings of foreign-born men. *J Polit Econ* 86:897–921.
39. Chiswick BR (1991) Speaking, reading, and earnings among low-skilled immigrants. *J Labor Econ* 9:149–170.
40. Rumberger RW, Larson KA (1998) Toward explaining differences in educational achievement among Mexican American language-minority students. *Sociol Educ* 71:69–92.
41. Dustmann C, Van Soest A (2002) Language and the earnings of immigrants. *Ind Labor Relat Rev* 55:473–492.
42. Berman E, Lang K, Siniver E (2003) Language-skill complementarity: Returns to immigrant language acquisition. *Labour Econ* 10:265–290.
43. Perreira KM, Harris KM, Lee D (2006) Making it in America: High school completion by immigrant and native youth. *Demography* 43:511–536.
44. van Ours JC, Veenman J (2006) Age at immigration and educational attainment of young immigrants. *Econ Lett* 90:310–316.
45. Beck A, Corak M, Tienda M (2012) Age at immigration and the adult attainments of child migrants to the United States. *Ann Am Acad Polit Soc Sci* 643:134–159.
46. Orrenius P, Zavodny M (2015) Does immigration affect whether US natives major in science and engineering? *J Labor Econ* 33:79–108.