Linguistic Traits and Human Capital Formation
(Forthcoming AEA Papers and Proceedings)

By Oded Galor, Ömer Özak and Assaf Sarid[*]

This research establishes the influence of linguistic traits on human behavior. Exploiting variations in the languages spoken by children of migrants with identical ancestral countries of origin, the analysis indicates that the presence of periphrastic future tense, and its association with long-term orientation has a significant positive impact on educational attainment, whereas the presence of sex-based grammatical gender, and its association with gender bias, has a significant adverse impact on female educational attainment.
JEL: D91, I25, J16, J24, Z10, Z13
Keywords: Human capital, Long-term Orientation, Gender Bias, Periphrastic Future Tense, Sex-Based Grammatical Gender, Culture, Language

The origins of the vast inequality in the wealth of nations have been attributed to the persistent effect of existing variations in the distribution of geographical, cultural, institutional, and human characteristics across the globe.[1] Variation in the evolution of linguistic traits in the course of human history has conceivably reinforced the persistent effect of cultural factors and contributed further to comparative economic development.

This research explores the impact of the coevolution of linguistic and cultural traits on the development process, shedding light on two fundamental questions: Has the evolution of linguistic traits contributed to the persistence of cultural characteristics and their lasting effect on economic prosperity? Have linguistic traits merely reflected existing cultural characteristics or have they influenced the human mindset and contributed directly to the development process?

In view of the pivotal role of languages in the transmission of knowledge, values, and beliefs, linguistic traits have plausibly reinforced the diffusion of cultural traits within and across generations. Thus, cultural characteristics that were embodied in language structures were more likely to persist across time and space, and natural selection favored linguistic traits that fortified prevailing cultural traits. Moreover, linguistic traits per se may have directly influenced the individual mindset and human behavior, beyond their indirect effect via the reinforcement of cultural transmission.

In particular, in societies characterized by distinct gender roles, and consequently by the existence of gender bias, a type of grammatical gender, which has plausibly fortified these cultural norms, emerged and persisted over time. Similarly, in societies characterized by long-term orientation, a structure of the future tense, which has presumably reinforced the efficiency of future oriented behavior, appeared and prevailed over time (Galor, Özak and Sarid, 2018).

Languages differ in the existence and the form of grammatical gender and the structure of the future tense. Languages that are characterized by the presence of *sex-based* grammatical gender classify nouns according to biological gender, inducing speakers to highlight gender distinctions. Furthermore, languages characterized by the ex-

[*] Galor: Department of Economics, Brown University; NBER, CEPR, IZA, CES-Ifo. E-mail: Oded_Galor@brown.edu. Özak: Department of Economics, Southern Methodist University; IZA. E-mail: ozak@smu.edu. Sarid: Department of Economics, University of Haifa. Email: asarid@econ.haifa.ac.il.
[1]Gallup, Sachs and Mellinger (1999), Guiso, Sapienza and Zingales (2006), Acemoglu, Johnson and Robinson (2001), Glaeser et al. (2004), and Ashraf and Galor (2013).

istence of *periphrastic* future tense, which is associated with discursive phrases such as 'will', 'shall', 'want to', and 'going to', reflect greater future intentions and orientation than the *inflectional* future tense that is associated with morphological variation in verbs (Bybee, Perkins and Pagliuca, 1994). Linguists as well as other scholars have therefore argued that gender biases have been reinforced by the presence of sex-based grammatical gender systems, while long-term oriented behavior have been fortified by the periphrastic future tense.

This research examines the effects of these two language structures on contemporary human capital formation, conceivably via their indirect effect on the persistence of ancestral cultural traits, as well as their direct effect on the individual mindset and human behavior. In particular, the analysis explores the impact of (i) the presence periphrastic future tense on college attendance, and (ii) the presence of sex-based grammatical gender on female college attendance.

The study advances a novel identification methodology that disentangles the impact of linguistic traits from the persistent effect of other ancestral characteristics, resolving major drawbacks that characterize existing explorations of the association between language structures and economic outcomes (Kashima and Kashima, 1998; Chen, 2013). In particular, this methodology augments the traditional epidemiological approach for the study of the persistent effects of culture by isolating the direct effect of linguistic traits from their potential indirect effects via their impact on the persistence of cultural traits.

I. Identification Strategy

The analysis exploits variations in language structures across individuals that are originated from the same ancestral homelands, transcending a major limitation of the influential epidemiological approach for the identification of the persistent effects of cultural traits on human behavior and economic outcomes (Giuliano, 2007; Fernandez and Fogli, 2009; Galor and Özak,

2016). While the epidemiological approach permits the exploration of the impact of the ancestral environment of migrants on their cultural traits, accounting for the geographical, institutional and cultural characteristics in the host country, it does not distinguish between the impact of cultural characteristics and other unobserved ancestral characteristics that reflect the parental countries of origin.

In contrast, the proposed methodology isolates the direct effect of linguistic traits on human behavior from the persistent effect of ancestral cultural characteristics. In particular, since children of migrants with identical ancestral countries of origin may speak different languages in the host country, one can disentangle the impact of linguistic traits on human behavior, from the impact of the ancestral environment, accounting for observed and unobserved geographical, institutional and cultural characteristics that characterized the ancestral homelands and may partly govern the behavior of children of migrants.

The analysis is conducted on individual data from the US Census and American Community Survey for the years 2000-2017 (Ruggles et al., 2019). It focuses on all children of migrants over the age of 24, who were either born in the US, or brought to the US before the age of 5. This sample consists of 747,062 individuals, whose parents migrated from 147 countries and speak 64 languages. The prevalence of sex-based grammatical gender and periphrastic future tense in these 64 languages is determined based on the classifications provided by Dryer (2013).

II. Periphrastic Future Tense and Education of Children of Migrants

This section explores the impact of speaking a language with periphrastic future tense on the probability of college attendance among children of migrants in the US.

Table 1 establishes the positive association between speaking a language with periphrastic future tense and college attendance. In particular, the estimates in

TABLE 1—PERIPHRASTIC FUTURE TENSE AND COLLEGE EDUCATION OF CHILDREN OF MIGRANTS

	College Attendance						
	All			Parental		No English	No Spanish
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Periphrastic Future Tense	0.232	0.226	0.053	0.038	0.035	0.053	0.029
	(0.057)	(0.057)	(0.020)	(0.012)	(0.012)	(0.024)	(0.015)
Geographical Controls (Language Homeland)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State & Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age, Gender, & Marital Status FE	No	Yes	Yes	Yes	Yes	Yes	Yes
Parental Country of Origin FE	No	No	Yes	Yes	Yes	Yes	Yes
Parental Education	No	No	No	Yes	Yes	Yes	Yes
Parental English Proficiency	No	No	No	No	Yes	Yes	Yes
R^2	0.05	0.13	0.16	0.21	0.21	0.23	0.24
Observations	735482	735482	735482	164722	164722	96738	96614

Notes: The table examines the impact of speaking a language with periphrastic future tense on the probability of college attendance among children of migrants in the US. Geographical characteristics in the historical homeland of the language include absolute latitude, mean elevation, mean ruggedness, coast length and pre-1500 crop return. Heteroskedasticity robust standard error estimates clustered at the country of origin, language and state levels are reported in parentheses.

columns (1)-(2) suggest that speaking a language with periphrastic future tense is associated with 23 percentage points increase in the probability of attending college, accounting for individual characteristics such as age, gender, marital status, state of residence, and year of interview, as well as the geographical characteristics of the historical homeland of the language. A sizable magnitude in comparison to a mean probability of 0.59 of college attendance in the sample as a whole. Column (3) further accounts for the parental countries of origin, namely, the ancestral geographical, institutional, and cultural characteristics that may affect individual human capital formation. Thus, the estimated association isolates the impact of long-term orientation that is language-embodied from the persistent cultural effects of long-term orientation via non-linguistic channels. The estimate implies that speaking a language with periphrastic future tense is associated with 5.3 percentage points increase in the probability of attending college.

Columns (4)-(5) establish that the estimated impact is unaffected qualitatively if one further accounts for parental education levels and their command of the English language. Not surprisingly, parental education and their level of proficiency in English have a positive and sizable association with their offspring's college attendance. Nevertheless, the estimates suggest that the impact of speaking a language with periphrastic future tense remains sizable and it is twice as large as the impact of having a parent that is proficient in the English language and nearly 1/3 of the impact of having a college educated parent. Finally, columns (6) and (7) establish that the results are unaffected qualitatively if English or Spanish speakers are excluded from the sample, accounting for augmented labor market opportunities and greater incentives to invest in human capital for individuals who are proficient in the two dominating languages in the US.

Thus, the analysis in Table 1 suggests that speaking a language with periphrastic future tense is associated with higher college attendance, accounting for a host of individual, socio-economic and ancestral characteristics.

III. Sex-Based Grammatical Gender and Education of Female Children of Migrants

This section explores the impact of languages characterized by the existence of sex-based grammatical gender on female human capital formation. In view of the proposed hypothesis that in a society characterized by distinct gender roles the existence of grammatical gender have reinforced prevailing gender biases, the analysis explores whether languages characterized by the existence of sex-based grammatical gender have an adverse impact on female human capital formation. In particular, following the identification strategy exploited in the previous section, the analysis focuses on the impact of sex-based grammatical gender on college attendance of female children of migrants in the US.

In line with the proposed hypothesis, Table 2 establishes a negative association between speaking a language with sex-based grammatical gender and female college attendance. In particular, columns (1) and (2) show that speaking a language with sex-based grammatical gender is associated with 23 percentage points lower probability of female college attendence, accounting for individual characteristics such as age, gender, marital status, state of residence, and year of interview, as well as the geographical characteristics of the historical homeland of the language. A sizable impact in comparison to a mean probability of 0.61 that a woman would attend college in the sample as a whole.

Nevertheless, this impact may capture the persistence of characteristics of the parental countries of origin of these women, independently of grammatical gender. Thus, column (3) accounts for parental origins fixed effects, and therefore isolates the impact of gender bias that is language-embodied from the persistent cultural effects of gender bias via non-linguistic channels. The results suggest that sex-based grammatical gender itself is associated with 5.5 percentage points lower probability of female college attendance.

Moreover, as established in columns (4) and (5), the adverse impact of speaking a language with sex-based grammatical gender on female college attendance is robust to the confounding effect of parental education and their proficiency in the English language. The estimated association between sex-based grammatical gender and the probability of attending college is sizable and it amounts to nearly 50% of the estimated impact of having a college educated parent. Finally, columns (6) and (7) show that excluding individuals who speak the two main languages in the US (i.e., English and Spanish) does not affect the qualitative results.

Thus, the analysis in Table 2 suggests that speaking a language with sex-based grammatical gender is negatively associated with female college attendance, accounting for a wide range of individual, socio-economic and ancestral characteristics.

IV. Concluding Remarks

This research establishes the influence of linguistic traits on human behavior. The analysis indicates that the presence of periphrastic future tense and its association with long-term orientation has a significant positive impact on educational attainment, whereas the presence of sex-based grammatical gender, and its association with gender bias, has a significant adverse impact on female educational attainment.

The impact of linguistic traits on contemporary human capital formation, may a priori operate indirectly, via their effect on the persistence of ancestral cultural traits, as well as directly via their effect on the individual mindset and human behavior. The study advances a novel identification methodology that plausibly disentangles the direct effect of linguistic traits on human behavior from its indirect effect via their impact on the persistence of ancestral cultural traits that may govern contemporary human behavior. In particular, it exploits variations in the language spoken by children of migrants with identical ancestral countries of origins to isolate the linguistic channel from the cultural one.

Table 2—Sex-Based Grammatical Gender and College Education of Female Children of Migrants

| | College Attendance | | | | | | |
| | All | | | Parental | | No English | No Spanish |
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Sex-Based Grammatical Gender	-0.240	-0.230	-0.055	-0.059	-0.053	-0.086	-0.046
	(0.065)	(0.060)	(0.024)	(0.019)	(0.020)	(0.047)	(0.018)
Geographical Controls (Language Homeland)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State & Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age, Gender, & Marital Status FE	No	Yes	Yes	Yes	Yes	Yes	Yes
Parental Country of Origin FE	No	No	Yes	Yes	Yes	Yes	Yes
Parental Education	No	No	No	Yes	Yes	Yes	Yes
Parental English Proficiency	No	No	No	No	Yes	Yes	Yes
R^2	0.05	0.13	0.16	0.20	0.20	0.20	0.25
Observations	345778	345778	345778	66267	66267	38323	34731

Notes: This table examines the impact of speaking a language with sex-based grammatical gender on female college attendance. Geographical characteristics in the historical homeland of the language include absolute latitude, mean elevation, mean ruggedness, coast length, average caloric suitability index and the average caloric yield of plow-negative crops. Heteroskedasticity robust standard error estimates clustered at the parental countries of origin, language and state levels are reported in parentheses.

REFERENCES

Acemoglu, Daron, Simon Johnson, and James A Robinson. 2001. "The Colonial Origins of Comparative Development: An Empirical Investigation." *The American Economic Review,* 91(5): 1369–1401.

Ashraf, Quamrul, and Oded Galor. 2013. "The out of Africa hypothesis, human genetic diversity, and comparative economic development." *The American Economic Review,* 103(1): 1–46.

Bybee, Joan L., Revere Perkins, and William Pagliuca. 1994. "The evolution of grammar."

Chen, M Keith. 2013. "The effect of language on economic behavior: Evidence from savings rates, health behaviors, and retirement assets." *The American Economic Review,* 103(2): 690–731.

Dryer, Matthew S & Haspelmath, Martin, ed. 2013. *The World Atlas of of Language Structures Online.* Leipzig: Max Planck Institute for Evolutionary Anthropology.

Fernandez, Raquel, and Alessandra Fogli. 2009. "Culture: An empirical investigation of beliefs, work, and fertility." *American Economic Journal: Macroeconomics,* 1(1): 146–177.

Gallup, John Luke, Jeffrey D Sachs, and Andrew D Mellinger. 1999. "Geography and economic development." *International regional science review,* 22(2): 179–232.

Galor, Oded, and Ömer Özak. 2016. "The Agricultural Origins of Time Preference." *American Economic Review,* 106(10): 3064–3103.

Galor, Oded, Ömer Özak, and Assaf Sarid. 2018. "Geographical Roots of the Coevolution of Cultural and Linguistic Traits." *NBER Working Paper w25289.*

Giuliano, Paola. 2007. "Living arrangements in western europe: Does cultural origin matter?" *Journal of the European Economic Association,* 5(5): 927–952.

Glaeser, Edward L, Rafael La Porta, Florencio Lopez-de Silanes, and Andrei Shleifer. 2004. "Do institutions cause growth?" *Journal of economic Growth,* 9(3): 271–303.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2006. "Does Culture Affect Economic Outcomes?" *Journal of Economic Perspectives*, 20(2): 23–48.

Kashima, Emiko S, and Yoshihisa Kashima. 1998. "Culture and language the case of cultural dimensionsand personal pronoun use." *Journal of Cross-Cultural Psychology*, 29(3): 461–486.

Ruggles, Steven, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas, and Matthew Sobek. 2019. *IPUMS USA: Version 9.0 [dataset].* IPUMS.

LINGUISTIC TRAITS AND HUMAN CAPITAL FORMATION

ODED GALOR, ÖMER ÖZAK AND ASSAF SARID

ONLINE APPENDIX

ROBUSTNESS TO VARIOUS SUBSAMPLES OF CHILDREN OF MIGRANTS

This appendix establishes the robustness of results to alternative sample of children of migrants: (i) migrant children born overseas who arrived to the US before the age of 5, ("one-and-a-half generation migrants"), (ii) children born in the US to at least one foreign parent ("second generation migrants"). The use of one-and-a-half and second-generation migrants overcomes a potential concern due to ethnic attrition bias (Duncan and Trejo, 2016). In particular, previous analyses that have employed the US census or ACS to study the effects of culture using migrants, have focused on all US-born individuals and tried to identify migrants and their ancestry by using individual's self-reported ancestry. Thus, these analyses have included all descendants of migrants that still identify with the country of origin of their ancestors. But, as Duncan and Trejo (2011, 2016), among others, have shown, individuals tend to self-identify differently depending on their generation, their true ancestry, and their socio-economic background. Thus, using second-and-higher-generation migrants can bias the results due to misidentification of ancestry. For this reason, the analysis is performed using one-and-a-half or second generation migrants. Robustness of the results to higher order migrants, as well as to other potential concerns, is established in Galor, Özak and Sarid (2016).

The sample of "one-and-a-half generation migrants," consists of 524,774 individuals, who migrated into the US before the age of 5. They were born in 147 countries and speak 64 languages.

The sample of "second-generation migrants" consists 222,288 offspring who were born in the US to at least one foreign born parent. These individuals originated from 143 countries of origin of the mother and 140 countries of origin of the father and they speak 63 languages.

	College Attendance							
	All			Parental			No English	No Spanish
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Periphrastic Future Tense	0.228	0.224	0.065	0.068	0.078	0.073	0.082	0.056
	(0.062)	(0.061)	(0.025)	(0.017)	(0.018)	(0.017)	(0.026)	(0.030)
Geographical Controls (Language Homeland)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State & Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age, Gender, & Marital Status FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Parental Country of Origin FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Parental Education	No	No	No	No	Yes	Yes	Yes	Yes
Parental English Proficiency	No	No	No	No	No	Yes	Yes	Yes
R^2	0.06	0.15	0.19	0.26	0.29	0.29	0.31	0.31
Observations	513028	513028	513028	30104	30104	30104	19664	17187

Notes: The table examines the impact of speaking a language with periphrastic future tense on the probability of college attendance among one-and-a-half generation migrants in the US. It replicates the analysis of Table 1 to show the robustness to this subsample. Geographical characteristics in the historical homeland of the language include absolute latitude, mean elevation, mean ruggedness, coast length and pre-1500 crop return. Heteroskedasticity robust standard error estimates clustered at the country of origin, language and state levels are reported in parentheses.

	College Attendance							
	All			Parental			No English	No Spanish
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Periphrastic Future Tense	0.224	0.221	0.027	0.026	0.027	0.025	0.047	0.027
	(0.054)	(0.052)	(0.011)	(0.010)	(0.010)	(0.010)	(0.024)	(0.013)
Geographical Controls (Language Homeland)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State & Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age, Gender, & Marital Status FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Parental Country of Origin FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Parental Education	No	No	No	No	Yes	Yes	Yes	Yes
Parental English Proficiency	No	No	No	No	No	Yes	Yes	Yes
R^2	0.05	0.13	0.17	0.18	0.21	0.21	0.22	0.24
Observations	214374	214374	214374	131057	131057	131057	74968	76206

Notes: The table examines the impact of speaking a language with periphrastic future tense on the probability of college attendance among second generation migrants in the US. It replicates the analysis of Table 1 to show the robustness to this subsample. Geographical characteristics in the historical homeland of the language include absolute latitude, mean elevation, mean ruggedness, coast length and pre-1500 crop return. Heteroskedasticity robust standard error estimates clustered at the country of origin, language and state levels are reported in parentheses.

	College Attendance						
	All			Parental		No English	No Spanish
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Periphrastic Future Tense	0.228	0.223	0.040	0.038	0.035	0.062	0.032
	(0.055)	(0.052)	(0.015)	(0.012)	(0.012)	(0.023)	(0.015)
Geographical Controls (Language Homeland)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State & Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age, Gender, & Marital Status FE	No	Yes	Yes	Yes	Yes	Yes	Yes
Maternal and Paternal Origin FE	No	No	Yes	Yes	Yes	Yes	Yes
Parental Education	No	No	No	Yes	Yes	Yes	Yes
Parental English Proficiency	No	No	No	No	Yes	Yes	Yes
R^2	0.06	0.14	0.18	0.22	0.22	0.23	0.24
Observations	130455	130455	130455	130455	130455	74709	75664

Notes: The table examines the impact of speaking a language with periphrastic future tense on the probability of college attendance among second generation migrants in the US. It replicates the analysis of Table 1 to show the robustness to this subsample. It extends the analysis of Table A2 by accounting simultaneously for the country of origin of both parents. Geographical characteristics in the historical homeland of the language include absolute latitude, mean elevation, mean ruggedness, coast length and pre-1500 crop return. Heteroskedasticity robust standard error estimates clustered at the countries of origin or both parents, language and state levels are reported in parentheses.

Table A4—Sex-Based Grammatical Gender and Female College Education: One-and-a-half Generation Migrants

| | College Attendance | | | | | | | |
| | All | | | Parental | | | No English | No Spanish |
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Existence of Sex-Based Gender System	-0.238	-0.233	-0.069	-0.063	-0.106	-0.096	-0.139	-0.086
	(0.067)	(0.061)	(0.025)	(0.030)	(0.032)	(0.033)	(0.067)	(0.043)
Geographical Controls (Language Homeland)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State & Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age, Gender, & Marital Status FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Marital Status FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Parental Country of Origin FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Parental Education	No	No	No	No	Yes	Yes	Yes	Yes
Parental English Proficiency	No	No	No	No	No	Yes	Yes	Yes
R^2	0.06	0.16	0.19	0.26	0.28	0.29	0.29	0.34
Observations	250910	250910	250910	11619	11619	11619	7425	5705

Notes: This table examines the impact of speaking a language with sex-based grammatical gender on college attendance among female one-and-a-half generation migrants in the US. It replicates the analysis of Table 2 to show the robustness to this subsample. Geographical characteristics in the historical homeland of the language include absolute latitude, mean elevation, mean ruggedness, coast length, average caloric suitability index and the average caloric yield of plow-negative crops. Heteroskedasticity robust standard error estimates clustered at the parental countries of origin, language and state levels are reported in parentheses.

TABLE A5—SEX-BASED GRAMMATICAL GENDER AND FEMALE COLLEGE EDUCATION OF SECOND GENERATION MI-
GRANTS

| | College Attendance | | | | | | | |
| | All | | | Parental | | | No English | No Spanish |
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Existence of Sex-Based Gender System	-0.209	-0.194	-0.012	-0.005	-0.033	-0.030	-0.061	-0.030
	(0.051)	(0.048)	(0.020)	(0.022)	(0.020)	(0.020)	(0.055)	(0.024)
Geographical Controls (Language Homeland)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State & Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age, Gender, & Marital Status FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Parental Country of Origin FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Parental Education	No	No	No	No	Yes	Yes	Yes	Yes
Parental English Proficiency	No	No	No	No	No	Yes	Yes	Yes
R^2	0.05	0.13	0.16	0.17	0.20	0.20	0.20	0.25
Observations	90660	90660	90660	52955	52955	52955	29978	27531

Notes: This table examines the impact of speaking a language with sex-based grammatical gender on
college attendance among female one-and-a-half generation migrants in the US. It replicates the analysis
of Table 2 to show the robustness to this subsample. Geographical characteristics in the historical homeland
of the language include absolute latitude, mean elevation, mean ruggedness, coast length, average caloric
suitability index and the average caloric yield of plow-negative crops. Heteroskedasticity robust standard
error estimates clustered at the parental countries of origin, language and state levels are reported in
parentheses.

	College Attendance						
	All			Parental		No English	No Spanish
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Existence of Sex-Based Gender System	-0.201	-0.181	-0.013	-0.038	-0.034	-0.064	-0.036
	(0.047)	(0.044)	(0.021)	(0.018)	(0.017)	(0.018)	(0.021)
Mom's Education Level (HS+)				0.123	0.125	0.112	0.117
				(0.010)	(0.010)	(0.015)	(0.015)
Dad's Education Level (HS+)				0.127	0.128	0.123	0.120
				(0.015)	(0.015)	(0.018)	(0.012)
Mom's English Level					0.014	0.017	-0.000
					(0.003)	(0.002)	(0.003)
Dad's English Level					0.008	0.009	0.009
					(0.002)	(0.001)	(0.003)
Geographical Controls (Language Homeland)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State & Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age, Gender, & Marital Status FE	No	Yes	Yes	Yes	Yes	Yes	Yes
Maternal and Paternal Origin FE	No	No	Yes	Yes	Yes	Yes	Yes
Parental Education	No	No	No	Yes	Yes	Yes	Yes
Parental English Proficiency	No	No	No	No	Yes	Yes	Yes
R^2	0.07	0.14	0.17	0.20	0.20	0.20	0.26
Observations	52734	52734	52734	52734	52734	29903	27339

Notes: This table examines the impact of speaking a language with sex-based grammatical gender on college attendance among female second migrants in the US. It replicates the analysis of Table 2 to show the robustness to this subsample. It extends the analysis of Table A5 by accounting simultaneously for the country of origin of both parents. Geographical characteristics in the historical homeland of the language include absolute latitude, mean elevation, mean ruggedness, coast length, average caloric suitability index and the average caloric yield of plow-negative crops. Heteroskedasticity robust standard error estimates clustered at the countries of origin or both parents, language and state levels are reported in parentheses.

*

Appendix References

Duncan, Brian, and Stephen J Trejo. 2011. "Intermarriage and the intergenerational transmission of ethnic identity and human capital for Mexican Americans." *Journal of Labor Economics*, 29(2): 195.

Duncan, Brian, and Stephen J Trejo. 2016. "The complexity of immigrant generations: Implications for assessing the socioeconomic integration of Hispanics and Asians." *NBER Working Paper Series*, , (w21982).

Galor, Oded, Ömer Özak, and Assaf Sarid. 2016. "Geographical Origins and Economic Consequences of Language Structures." Institute for the Study of Labor (IZA).