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.

Table A3—Periphrastic Future Tense and College Education of Second Generation Migrants

| | 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.

8

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 MIGRANTS

| | 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.

11

*

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).