



# The Impact of Linguistic Distance from English on Economic Growth: A Cross-Country Analysis

Zeynep Ozkok\*, Brandon Malloy, Amy Rowe

Department of Economics, St. Francis Xavier University, Antigonish, Canada. \*Email: [zozkok@stfx.ca](mailto:zozkok@stfx.ca)

Received: 06 December 2021

Accepted: 05 February 2022

DOI: <https://doi.org/10.32479/ijefi.12725>

## ABSTRACT

Increasing levels of globalization have contributed to English becoming a preferred language of international communication, education, and trade. With countries rapidly embedding English into a large variety of curricula, the demand for English as a medium of instruction is ever growing. Using information from the US Foreign Service Institute's Professional Working Proficiency values, we identify seven distinct linguistic distance from English (LDE) categories and examine the effects of linguistic distance on income and economic growth across 97 countries over the 1980-2018 period. In addition to the direct effects of linguistic distance on incomes and growth, we further explore its impact through the channels of education and trade. Our results demonstrate that linguistic distance from English affects income levels and growth non-linearly across countries- increased distance from English does not necessarily translate into declining levels of income or growth rates. Conversely, we find that, via the trade channel, more distant languages tend to experience positive rates of growth. Controlling for regional effects or the existence of multiple languages of instruction do not alter our findings. Sub-dividing our dataset by national income levels, we show that the largest negative effects on growth are among Upper-Middle Income countries.

**Keywords:** Linguistic Distance, English, Language, Income and Economic Growth, Trade, Education

**JEL Classifications:** Z10, F10, I20, O47

## 1. INTRODUCTION

With rising levels of globalization, English has become the preferred language of instruction and communication across the world. Recent statistics show that more than a billion people around the world now speak English as a first or second language (Education First, 2020). The English language has played a major role in upward social mobility and is considered a prerequisite for scientific and technological development in many countries (Ndamba et al., 2017; Rubagumya, 1989). Since the start of the twenty first century, English has increasingly been employed across institutions, with countries rapidly embedding English into a large variety of curricula, through extensive foreign language teaching, or more prominently as the medium of instruction in schools. This is particularly noticeable in countries where English is not the official language (Marsh, 2006; Rose and McKinley,

2016). Nevertheless, language acquisition costs are often significant, and vary widely across countries with different native languages. For example, it is commonly understood that speakers of Romance languages like French or Italian find it easier to learn English (and vice-versa) due to common root words, conjugation, and grammatical structures, in comparison to speakers of Sino-Tibetan or Afro-Asiatic languages. Yet despite widespread and increasing adoption of the English language in official settings and as a medium of instruction, the benefits of English language usage on economic growth remain unclear. Does English language usage generate additional gains, in excess of educational attainment, on economic growth across countries? Does the magnitude of the "distance-from-English" of languages used across countries impact economic growth rates, and if so, in what direction? Through which channels does linguistic distance from English (LDE) influence incomes and economic growth? In this paper, we explore these

questions by examining the effect of English language usage, and the linguistic distance from English of various languages on income levels and economic growth across countries.

Our paper contributes to two distinct branches of the literature. The first examines the challenges of developing quantitative measures of linguistic diversity within and across countries and assesses their impact on economic growth, welfare, trade, and migration. Desmet et al. (2009) show that linguistic diversity across countries has a statistically significant effect on redistribution, economic growth, and public goods. Goren (2018), analyzing the link between ethno-linguistic diversity and economic growth in a panel of developed and developing countries from 1960-2009, finds distance-weighted diversity measures to be negatively related to economic growth. In a study using Levenshtein distances to create a measure of linguistic distance, Isphording and Otten (2013) find higher language acquisition costs for immigrants from countries whose language is more distant from the host country's. In quantifying the disadvantage in literacy skills arising from linguistic distances between immigrant mother tongues and host country languages, Isphording (2014) shows linguistic distance leads to a clear disadvantage in literacy scores among immigrants, making the assimilation process of immigrants more difficult. Alternatively, a common language, or to a lesser extent a similar or related language, is found to reduce transaction costs and increase trade between countries (Isphording and Otten, 2013). Melitz (2008) demonstrates that a common language across countries is important in promoting bilateral trade through ease of communication and translation. Similarly, examining 71309 pairs of FDI relationships between 2000 and 2012, Ly et al. (2018) find a positive impact of language on the level of FDI in various countries. In studying the effect of linguistic distances on migration, using data from 30 OECD countries over the 1980-2010 period, Adsera and Pytlikova (2015) show migration rates to be increasing with linguistic proximity, with similar languages amounting to 14% to 20% of a rise in migration.

The second strand of literature focuses on the differences in linguistic distance from the English language or English proficiency and their effects on immigration and international trade. Examining the importance of linguistic distance on bilateral trade with the US for a sample of 36 non-English speaking countries, Hutchinson (2005) finds linguistic distance from English to have a significant negative effect on international trade, with the volume of imports decreasing in linguistic distance from English. In a study of the effect of English language proficiency on bilateral trade flows, Ku and Zussman (2010) show English proficiency to have a positive and significant effect on bilateral trade, stressing the importance of English proficiency in helping countries overcome barriers of language in trade. In trying to capture the distance between languages, Chiswick and Miller (2005) use the linguistic distance from English to analyze the proficiency in English among immigrants in the US and Canada. The results demonstrate that a greater distance between an immigrant's native tongue and English leads to a lower level of the immigrant's English linguistic proficiency. In a later study, Chiswick and Miller (2012) find that the positive assimilation of immigrants into the US depends on the linguistic distance of the immigrants' mother tongues from

English. The results show earnings among the immigrants to be higher for those whose mother tongue is closer to English.

To address the relative paucity of information focusing on the quantitative effect of linguistic distance from English on incomes and growth across countries and time, we construct a dataset identifying the medium of instruction in primary and secondary educational institutions across 97 countries. We use information from the US Foreign Service Institute's School of Language Studies (FSI-SLS) to create a quantitative measure of the linguistic distance of these languages from English, based on the learning "timeline *usually* required for a student to reach "Professional Working Proficiency" in the language".<sup>1</sup> Using this source to construct our linguistic distance measure is particularly novel, as it allows us to include a broader set of languages in our sample (over 65 distinct languages) than is typically employed in cross-country analyses. Furthermore, to account for any potential country-specific effects, we group our linguistic distance measures into seven categories, which we believe offers an improvement to previously analyzed linguistic distance from English (LDE) measures. Our empirical analysis then examines the effect of this linguistic distance measure on average income levels- measured by log GDP per capita- and economic growth rates- measured by the annualized percentage growth rate in GDP per capita- across countries between 1980 and 2018. We further decompose our analysis by investigating the interactions between our linguistic distance measures and education and trade variables to better understand the link between linguistic distance from income and growth through these specific channels.

We find two categories of languages, LD2 (German) and LD7 (Japanese) to have positively significant effects on income levels, but negatively significant effects on growth rates. Other categories of linguistic distance from English are mainly found to have less significant effects on incomes and growth. The inclusion of a dummy variable to capture multiple languages of instruction and the addition of regional dummies into our empirical analysis do not alter our findings from the benchmark model. Our consideration of two channels through which linguistic distance from English could impact incomes and growth rates similarly shows a positive effect of LD2 and LD7 categories on incomes and a negative effect of the same categories on growth rates. Under both channels, we find a non-linear and discontinuous effect of linguistic distance measures on our dependent variables. The impact of linguistic distance from English does not increase linearly or continuously across language categories. An examination of the interaction terms of linguistic distance with education and trade display higher significance levels via the trade channel. When considering the partial effects with respect to education, languages least distant from English are found to have a positive effect on incomes and growth rates. In the trade channel, partial effects display positive effects on growth for languages most distant from English. The total effects are found to be positive across five categories of languages for income levels, while for economic growth, we mainly find a negative effect. One intriguing result is the positive

<sup>1</sup> For a more detailed description, visit <https://www.state.gov/foreign-language-training/>

effect of the higher LD categories (such as LD7) on growth rates- to further delve into possible explanations for this, decompose our analysis across four national income-level groupings. Our results suggest that languages least distant from English have the largest positive effects on incomes in Low Income, Upper-Middle Income and High Income countries, while the largest negative effects on growth are found among Upper-Middle Income countries.

Our findings contribute to the literature by introducing a novel measure of linguistic distance from English that incorporates both a larger number of potential languages and countries, and allows for the examination of the impact of these measures on income and growth. Moreover, with increasing levels of linkages across countries due to international trade and globalization, and the heightened demand for English medium of instruction on the education side, we analyze the trade and education channels through which linguistic distance could additionally affect income and growth.

The remainder of our paper is structured as follows. In the next section, we describe the data in detail. Section 3 presents the empirical model and the results while Section 4 offers a review of robustness checks. Section 5 concludes.

## 2. DATA

The data used in this study come from a variety of sources, resulting in a panel covering 97 countries during the period from 1980 to 2018. Our main variables of interest, the linguistic distance from English measures, are constructed using information obtained from the US Foreign Service Institute's *School of Language Studies (FSI-SLS)* website. Our dependent variables of interest- per capita income and economic growth across countries- and our control variables are obtained predominantly from the World Bank's *World Development Indicators* database and the University of Gothenburg's *Quality of Government* database. In this section, we discuss the indicators employed in our analysis in further detail.

### 2.1. Measures of Linguistic Distance from English

To construct our measures of linguistic distance from English, we first compile a list of medium-of-instruction (MOI) languages used in primary and secondary education, as reported by each country's Ministry of Education websites or equivalent official sources. Most countries use the same languages for both primary and secondary education. In countries with various regional languages, it is common for a local language to be used as an additional or alternative medium of instruction at the primary level, alongside a more widely used language at the country level. For those countries, we take the more widely used language as the MOI, and include a dummy variable (MultMOI) which takes on a value of 1 in countries that have multiple MOI languages, and a value of 0 otherwise. However, in these instances, it is common for both, or all languages used to fall into the same linguistic distance category.<sup>2</sup> We also create a dummy variable (Engl=1) to

track whether or not English is listed as one of the languages used in primary and/or secondary education in countries with multiple languages of medium of instruction.<sup>3</sup>

Having identified the prevailing MOI languages for each country, we then construct our linguistic distance from English measures using information provided by the FSI-SLS's "Professional Working Proficiency" index. The FSI's website offers a timeline for language learning, specifying the average length of time (measured in weeks or class-hours), needed for a native English speaker (among US diplomatic staff) to achieve "Professional Working Proficiency" in a given language, which corresponds to an average grade of "Speaking-3, Reading-3" on the Interagency Language Roundtable scale.<sup>4</sup> One particular advantage of this approach is that it allows us to quantify the relative difficulty of language acquisition for a wider variety of languages than are typically included in these types of analyses.<sup>5</sup> In total, the FSI lists 67 languages, which are grouped into four main categories. We extend the FSI classification into the following seven categories:<sup>6</sup>

- Category 0: English language. This group includes all countries that use the English language as a primary and/or secondary MOI and serves as our base case for our quantitative analysis.
- Category 1: Languages that are "closely related to English" and take approximately 23-24 weeks (575-600 class hours) to achieve professional working proficiency. Examples include Danish, Italian, Spanish, Dutch, French and Portuguese.
- Category 2: Languages that are "similar to English" and take approximately 30 weeks (750 class hours) to achieve professional working proficiency. This category is exclusively composed of the German language.
- Category 3: Languages with "linguistic and/or cultural differences from English" and take 36 weeks (900 class hours) to achieve professional working proficiency. Languages included are Haitian Creole, Malay, Swahili and Indonesian.
- Category 4: Languages with "significant linguistic and/or cultural differences from English" and take approximately 44 weeks (1100 class hours) to achieve professional working proficiency. Some examples include Greek, Persian, Turkish, and Russian.

---

Category 1 languages, Belgium is classified as a Category 1 country. Other examples include Luxembourg, South Africa and Switzerland.

- 3 For example, Switzerland uses multiple medium-of-instruction languages, such as German, French and Italian, but not English, thus MultMOI=1 and Engl=0 for Switzerland; conversely, Canada uses both English and French, thus MultMOI=1 and Engl=1 for Canada. Finally, the United Kingdom uses only English as a method of instruction, thus MultMOI=0 and Engl=1 for the UK
- 4 For a more detailed explanation of the Interagency Language Roundtable's requirements for Speaking-3 and Reading-3 level proficiency, visit <https://www.govtirl.org/Skills/ILRscale2.htm> and <https://www.govtirl.org/Skills/ILRscale4.htm> respectively.
- 5 For more information on this earlier FSI categorization please visit <https://bigthink.com/strange-maps/how-long-to-learn-that-language-heres-a-map-for-that/>
- 6 Our classification of categories is based on an amendment of an earlier categorization by the FSI. We augment this classification by differentiating two languages, German and Japanese, into their own categories, as well as generating a separate category for some languages within "Category 4" languages deemed to be relatively more difficult to learn by the FSI than other languages in this group.

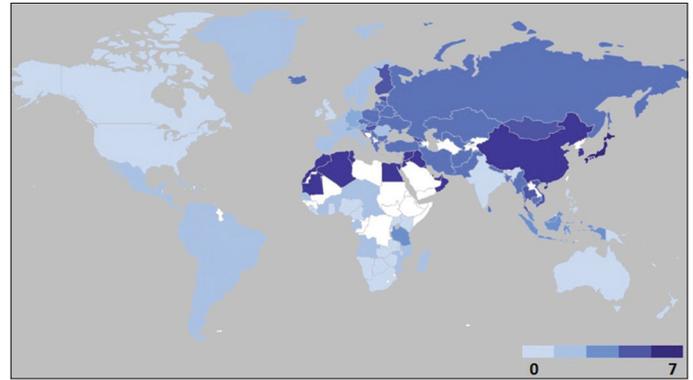
2 For example, in Belgium, Dutch, French and German are all used as a MOI in various primary and secondary schools. Given that both Dutch and French are both prominent languages used, and they both fall under

- Category 5: Languages that are deemed more difficult than those listed under Category 4 by the FSI. These languages may take longer than 44 weeks (1100 class hours) to achieve professional working proficiency. Some of the languages included in this category are Estonian, Finnish and Thai.
- Category 6: Languages which are “exceptionally difficult for native English speakers” to learn and take approximately 88 weeks (2200 class hours) to achieve professional working proficiency. Languages in this category include Arabic, Cantonese, Mandarin and Korean.
- Category 7: Languages, which are “additionally” more difficult than those in Category 6. This category is composed exclusively of Japanese.

This allows us to create the seven categories of languages listed above and construct an ordinal measure for the linguistic distance from English in order of increasing difficulty: languages with the most linguistic similarities to English are assigned a value of one (Linguistic Distance=1 or LD1), while languages with the least similarities are assigned a value of seven (LD7).

Table 1 presents our complete ordering of linguistic distance from English for all languages included in our dataset. One key assumption in the construction of these measures of linguistic distance from English is symmetry- that the average length of time for learning a foreign language by an English speaker is roughly equivalent to the time it takes for a native speaker of that language to learn English. While this may not hold in a strictly cardinal

**Figure 1:** Linguistic distance from English, by Country



**Table 1: Languages by Linguistic Distances from English**

LD0	LD1	LD2	LD3	LD4	LD5	LD6	LD7
English	Danish	German	Haitian Creole	Albanian	Estonian	Arabic	Japanese
	Dutch		Indonesian	Amharic	Finnish	Chinese-Cantonese	
	French		Malay	Armenian	Georgian	Chinese- Mandarin	
	Italian		Swahili	Azerbaijani	Hungarian	Korean	
	Norwegian			Bengali	Mongolian		
	Portuguese			Bulgarian	Thai		
	Romanian			Burmese	Vietnamese		
	Spanish			Czech			
	Swedish			Dari			
				Farsi			
				Greek			
				Hebrew			
				Hindi			
				Icelandic			
				Kazakh			
				Khmer			
				Kurdish			
				Kyrgyz			
				Lao			
				Latvian			
				Lithuanian			
				Macedonian			
				Nepali			
				Polish			
				Russian			
				Serbo-Croatian			
				Sinhala			
				Slovak			
				Slovenian			
				Somali			
				Tagalog			
				Tajiki			
				Tamil			
				Telugu			
				Tibetan			
				Turkish			
				Turkmen			
				Ukrainian			
				Urdu			
				Uzbek			

**Table 2: Countries by Linguistic Distances from English**

LD0	LD1	LD2	LD3	LD4	LD5	LD6	LD7
Australia	Argentina	Austria	Indonesia	Albania	Estonia	Algeria	Japan
Bahamas	Belgium	Germany	Malaysia	Armenia	Finland	Bahrain	
Barbados	Benin	Switzerland	Tanzania	Bangladesh	Georgia	Chad	
Belize	Bolivia			Belarus	Hungary	China	
Botswana	Brazil			Bulgaria	Mongolia	Egypt	
Cameroon	Burkina Faso			Cambodia	Thailand	Iraq	
Canada	Central African Republic			Croatia	Vietnam	Jordan	
Fiji	Chile			Cyprus		South Korea	
The Gambia	Colombia			Czech Republic		Kuwait	
Ghana	Costa Rica			Greece		Lebanon	
Grenada	Cote d'Ivoire			Iceland		Mauritania	
India	Cuba			Iran		Morocco	
Ireland	Denmark			Israel		Oman	
Jamaica	Dominican Republic			Kazakhstan		Qatar	
Kenya	Ecuador			Latvia		Syria	
Liberia	El Salvador			Lithuania		Tunisia	
Malawi	France			Myanmar		United Arab Emirates	
Malta	Gabon			Nepal			
Mauritius	Guatemala			Pakistan			
Namibia	Honduras			Poland			
New Zealand	Italy			Russia			
Nigeria	Luxembourg			Slovakia			
Philippines	Madagascar			Slovenia			
Rwanda	Mexico			Sri Lanka			
Sierra Leone	Moldova			Turkey			
South Africa	Mozambique			Ukraine			
Trinidad and Tobago	Netherlands			Uzbekistan			
Uganda	Nicaragua						
United Kingdom	Niger						
United States	Norway						
Zambia	Panama						
Zimbabwe	Paraguay						
	Peru						
	Portugal						
	Senegal						
	Spain						
	Suriname						
	Sweden						
	Togo						
	Uruguay						
	Venezuela						
(32)	(41)	(3)	(3)	(27)	(7)	(16)	(1)

sense, we believe that the ordering holds when inverted, and thus our ranking accurately reflects the relative difficulty for non-native English speakers to become proficient in English. Table 2 lists the countries included in our analysis, sorted according to their linguistic distance categories.

Figure 1 provides a geographic heatmap visualization of our linguistic distance measures for the countries included in our dataset. As we observe some evidence of geographical correlation of LDE values across countries, we consider adding regional controls to our quantitative analysis.<sup>7</sup>

Our linguistic distance from English measure differs notably from those used in the literature. The most similar measure used is that of Chiswick and Miller (2005), which is a scalar

<sup>7</sup> Some countries in our geographic heatmap appear in white. These are countries for which we do not have data. As a result, we do not include the countries displayed in white in our empirical estimations.

measure of the distance between English and other languages. This measure is computed using various language scores reported by Hart-Gonzalez and Lindemann (1993), which focuses on the average exam scores after 24 weeks of lessons in speaking proficiency by English-Speaking Americans at the US Department of State's School of Language Studies (Chiswick and Miller, 2015). This measure is one that groups languages into three different categories, assigning languages a linguistic score between 1 to 3, with scores increasing with similarity to the English language (Chiswick and Miller, 2005). However, in contrast to the relatively limited scope of their measure, our linguistic distance measure is able to capture a larger and more diverse group of languages according to their level of difficulty as established by the FSI. With the wide list of languages provided by the FSI, we believe that our linguistic distance measure from English offers a more in-depth examination of linguistic differences across countries.

## 2.2. Measures of Income and Growth

We use two dependent variables in our quantitative analysis. To measure the “level” of average incomes across countries, we use GDP per capita (measured in constant 2010 US dollars), which we then log-transform to account for potential heteroskedasticity issues. To measure “growth” in average incomes, we use growth in GDP per capita, measured in percentage terms. Both variables are obtained directly from the World Bank’s World Development Indicators.

## 2.3. Control Variables

We include the gross secondary school enrollment rate (in percentage terms) as a measure of differences in education across countries to investigate the impact of linguistic distance from English on incomes and growth through the education channel.<sup>8</sup> Similarly, to examine the impact of linguistic distance through the channel of trade, we include trade openness, as measured by the sum of imports and exports, as a share of GDP. We also use foreign direct investment (FDI), net national investment (in non-financial assets), and government expenditure (in constant 2010 US dollars) in our empirical estimations. These variables help control for the effects of increasing levels of financial globalization, as well as investment and government spending rates across countries respectively. Two employment variables- employment in industry and employment in services (both in percentages of total employment) - are included to control for the sectoral shifts of labour markets across countries.<sup>9</sup> To account for demographic shifts over time, we use both the population growth rate (in percentage terms) and the urban population share.

To examine such a broad swath of countries, particularly many low- and middle-income countries, we consider controls

- <sup>8</sup> Admittedly, a measure of educational attainment, such as performance on standardized testing, would be a preferable measure of human capital accumulation through education, affecting incomes and growth. However, these measures are often lacking or unavailable for lower income countries. In order to avoid restricting our analysis to a much smaller subset of countries, we use the more widely available metric of secondary enrollment rate as a measure of educational participation across countries.
- <sup>9</sup> The baseline is therefore employment in agriculture, which consists of the remaining share of employment in the labour market.

for institutional quality that are obtained from the Quality of Government database from the University of Gothenburg. We include a measure for the “level of democracy”- a variable on a 0-10 scale (with higher values corresponding to greater democracy), as calculated by the Freedom House index (Teorell et al., 2021). Alternatively, we use a control for the level of corruption in institutions, derived from the International Country Risk Guide by the PRS Group. This variable helps measure corruption, law and order, and the quality of bureaucracy and takes on a value between 0 and 1, where higher values represent higher risk of institutional corruption (Teorell et al., 2021).

Lastly, we include regional controls to account for differences in income levels across countries. Regional dummies are constructed using the modified versions of the regions classified in the World Bank analytical grouping. Categories include East Asia & Pacific, Eastern Europe, Latin America & Caribbean, Middle East & North Africa, South America, South Asia, and Sub-Saharan Africa. We also add an OECD dummy to capture the potential positive effects of being a high-income OECD member country. Similarly, an OPEC dummy is constructed to account for the potential positive effects of increased country wealth due to oil reserves and oil production. Table 3 presents the summary statistics for all the variables included in our analysis.

## 3. EMPIRICAL SPECIFICATION AND RESULTS

Our empirical analysis employs a model that measures the effects of linguistic distance from English on income levels and economic growth. The benchmark model, using a random effects estimation given below, controls for average differences across countries and time. It includes our seven-category linguistic distance measure (LD1-LD7, with LD0 representing English as the base group), our macroeconomic and institutional control variables and regional dummy variables. The two models estimated under our benchmark model are as follows:

**Table 3: Summary Statistics**

Variable	Mean	Std. Dev.	Min	Max	Obs
Log GDP per capita (Level)	8.47	1.54	4.90	11.66	4666
GDP per capita growth (Growth)	1.79	5.16	-64.99	53.97	4674
Secondary Enrollment Rate (SER)	73.24	31.64	2.48	163.93	3792
Trade Openness	75.42	42.62	0.02	416.39	4520
Foreign Direct Investment (FDI)	3.68	14.33	-58.32	451.64	4566
Government Expenditure	15.76	5.86	0.91	76.22	4426
Net Investment	2.54	2.36	-7.98	29.51	2609
Share of Employment: Industry	21.07	8.47	1.70	59.58	3510
Share of Employment: Services	51.78	18.06	7.23	87.85	3510
Democracy Index	6.39	3.26	0.25	10.00	4660
Corruption Index	0.43	0.22	0.00	0.96	3766
Urban Population	56.95	22.32	4.72	100.00	4978
Population Growth Rate	1.57	1.56	-9.08	17.51	4973
OPEC	0.08	0.26	0	1	4978
OECD	0.26	0.44	0	1	4978

FDI, Government Expenditure, and Net Investment are measured as the percentage of GDP; GDP per capita growth, Secondary Enrollment Rate, Urban Population, Population Growth Rate are measured in percentages; Democracy Index: 0 to 10, where higher values represent more democracy; Corruption Index: 0 to 1, where higher values demonstrate a rise in the level of corruption.

**Table 4: Baseline Specifications**

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	Level	Growth	Level	Growth	Level	Growth
LD1	-0.139 (0.167)	-0.758 (0.496)	-0.134 (0.168)	-0.730 (0.494)	-0.148 (0.168)	-0.476 (0.489)
LD2	0.566*** (0.170)	-2.090*** (0.699)	0.556*** (0.163)	-2.206*** (0.826)	0.404*** (0.141)	-1.900** (0.743)
LD3	0.206 (0.185)	-1.155 (0.864)	0.194 (0.177)	-1.325 (0.951)	0.226 (0.197)	-1.702* (0.880)
LD4	0.176 (0.179)	-0.824 (0.658)	0.176 (0.178)	-0.846 (0.635)	-0.131 (0.145)	-0.319 (0.571)
LD5	-0.0734 (0.320)	-1.803*** (0.668)	-0.0754 (0.317)	-1.847*** (0.634)	-0.141 (0.289)	-1.305** (0.527)
LD6	0.277 (0.297)	0.244 (0.922)	0.269 (0.293)	0.288 (0.893)	-0.212 (0.277)	1.227 (0.804)
LD7	0.398*** (0.154)	-1.817*** (0.537)	0.410** (0.160)	-1.772*** (0.530)	0.700* (0.373)	-1.395* (0.784)
MultMOI			0.0648 (0.136)	0.618 (0.450)		
TradeOpen	0.00156*** (0.000485)	0.0128*** (0.00343)	0.00155*** (0.000489)	0.0120*** (0.00370)	0.00153*** (0.000490)	0.0122*** (0.00340)
FDI	0.000240** (9.63e-05)	-0.00369 (0.00534)	0.000241** (9.64e-05)	-0.00348 (0.00536)	0.000241** (0.000100)	-0.00432 (0.00527)
SecEnroll	0.00178** (0.000905)	0.00534 (0.0103)	0.00178** (0.000905)	0.00406 (0.0105)	0.00187** (0.000912)	0.00814 (0.0102)
GovtExpGDP	-0.0159*** (0.00391)	-0.251*** (0.0498)	-0.0159*** (0.00392)	-0.252*** (0.0496)	-0.0159*** (0.00386)	-0.280*** (0.0543)
NetInvest	0.00761 (0.00572)	0.162* (0.0874)	0.00762 (0.00573)	0.169* (0.0897)	0.00754 (0.00587)	0.114 (0.0860)
EmpIndustry	0.0138** (0.00690)	0.0317 (0.0275)	0.0137** (0.00690)	0.0334 (0.0276)	0.0130* (0.00685)	0.0546** (0.0267)
EmpServices	0.0294*** (0.00462)	-0.0779*** (0.0283)	0.0294*** (0.00462)	-0.0784*** (0.0280)	0.0290*** (0.00458)	-0.0642** (0.0278)
Democracy	0.0273*** (0.00928)	-0.00699 (0.0827)	0.0274*** (0.00931)	0.0163 (0.0899)	0.0274*** (0.00926)	0.0225 (0.0857)
Corrupt	-0.0533 (0.0948)	-1.323 (1.127)	-0.0519 (0.0951)	-1.137 (1.157)	-0.0240 (0.0968)	-1.751 (1.191)
UrbanPop	0.0136*** (0.00308)	0.0169 (0.0179)	0.0136*** (0.00308)	0.0169 (0.0173)	0.0126*** (0.00320)	0.0288 (0.0200)
PopGR	0.0256** (0.0123)	-1.081*** (0.222)	0.0255** (0.0123)	-1.111*** (0.212)	0.0261** (0.0125)	-1.174*** (0.240)
OPEC	0.110 (0.0763)	-1.124 (0.717)	0.110 (0.0767)	-0.914 (0.770)	0.0881 (0.0746)	-0.773 (0.593)
OECD	1.283*** (0.166)	0.299 (0.510)	1.282*** (0.166)	0.331 (0.506)	0.826*** (0.155)	0.284 (0.482)
Constant	5.288*** (0.351)	9.494*** (1.570)	5.279*** (0.354)	9.315*** (1.615)	6.012*** (0.430)	7.016*** (1.602)
Observations	1,583	1,582	1,583	1,582	1,583	1,582
Region Dummies	N	N	N	N	Y	Y
Number of Countries	97	97	97	97	97	97

Robust standard errors in parentheses

\*\*\* p&lt;0.01, \*\* p&lt;0.05, \* p&lt;0.1

$$\text{LogGDPpc}_{j,t} = \alpha + \sum_{n=1}^7 \delta_n \text{LD}(n)_j + \sum_k \beta_k X_{k,j,t} + u_j + \epsilon_{j,t} \quad (1)$$

$$\text{GDPpcGrowth}_{j,t} = \alpha + \sum_{n=1}^7 \delta_n \text{LD}(n)_j + \sum_k \beta_k X_{k,j,t} + u_j + \epsilon_{j,t} \quad (2)$$

where for each country  $j$ , in period  $t$ ,  $X_k$  represents our  $k \times 1$  vector of control variables  $u_j$  are country-specific random effects, and our coefficients of interest are therefore:  $\delta_1, \delta_2, \delta_3, \delta_4, \delta_5, \delta_6$ . Given the nature of our unbalanced panel data, we cluster our standard

errors at the country-level. Because of the data limitations for some of our control variables, we typically retain 97 countries in our regression analyses, with nearly 1600 country-year observations for most of our specifications.

### 3.1. Baseline Specification

In Table 4, columns (1) and (2) present the results of our benchmark model, using our seven linguistic distance variables of interest, as well as our macroeconomic controls. We find two of the seven linguistic distance coefficients to be positively significant for income, and three of the seven coefficients to be negatively significant for GDP growth. More specifically, our results suggest

that countries where the medium of instruction is in group LD2 category of languages (corresponding to the German language) exhibit a higher level of mean incomes, but a lower rate of annual growth in GDP per capita, when compared to countries where the MOI is English (LD0). We observe the same pattern with LD7 (corresponding to the Japanese language). This result may be reflective of historical, cultural, or economic idiosyncrasies specific to Germany (and/or other German speaking countries) and to Japan, which in our analysis are being captured by their respective linguistic distance coefficient estimates. Additionally, we find the LD5 category of languages to have a significant negative effect on growth. Given the inclusion of EU countries such as Finland, Hungary, and Estonia in this group, we believe that the negative effect could be driven by the relative difficulty of learning these languages as well as the lower potential of yearly GDP growth.

In our baseline specification, most of our macroeconomic controls exhibit the expected signs, consistent with the preponderance of growth literature. We see increasing levels of mean income in countries with more trade openness, higher FDI, higher secondary education enrollment rates, lower government expenditure ratios, higher population growth rates, and higher rates of urbanization. We observe higher levels of income in countries with a higher democracy score. We find evidence of higher income levels among countries who have moved from agriculture-based economies to manufacturing, and further into service-oriented economies.

Turning to growth rates, we find higher growth rates among countries with more trade openness, and lower growth rates in countries with higher shares of government expenditure. Net investment is also found to have a positive effect on growth rates across countries. Consistent with neoclassical growth theories, we find lower growth rates among countries with higher population growth rates. We also observe lower growth rates among countries where sectoral shifts have led to a higher proportion of production in service industries, when compared to manufacturing and agriculture.

In columns (3)-(4) of Table 4, we add a control for countries who use multiple languages of instruction, in addition to our seven linguistic distance measures. However, our Multiple-MOI dummy is not statistically significant in either the Level or the Growth regressions, and the inclusion of this variable does not significantly alter our other estimates.<sup>10</sup>

In columns (5)-(6), we include regional dummy variables to control for potential geographic differences that influence incomes and growth. The inclusion of these regional controls does not alter the coefficient estimates of our linguistic distance measures greatly, compared to our baseline specification; however, it does result in our coefficient estimate for the LD3 category (including languages such as Haitian Creole, Indonesian and Malay) becoming statistically significant and negative, although only at a 90% confidence level.

<sup>10</sup> We find similar results when we include our English dummy variable, which we do not report in our table, for brevity. Due to this lack of statistical significance, we do not include these variables in the remainder of our empirical estimations. The results with the inclusion of the English dummy are available upon request.

### 3.2. Baseline with Interaction Terms

Our baseline specification provides some muted evidence for an impact of linguistic distance from English on income levels or growth over time for certain language categories. However, the relationship is not linear- to better understand how linguistic distance income and growth rates across countries, we also investigate whether our linguistic distance measures have an indirect impact on our dependent variables, via the channels of education and trade. To address this, we add interaction terms for each of our seven linguistic distance variables with our trade openness variable and our secondary enrollment rate variable. This helps determine whether variation in linguistic distance from English across countries may alter the returns to trade and education on income levels and/or growth.

Table 5 reports the results of multiple regressions in which we add the interaction terms to our baseline specification. The first two columns represent the results from our baseline specification when we include education interaction terms using our education variable, *SecEnroll*, while in columns (3)-(4) we add trade interaction terms using our trade variable, *TradeOpen*, and finally, in columns (5)-(6) we include both sets of interaction terms simultaneously. We find similar results to the baseline specification with the inclusion of education interaction terms on income as shown in column (1). The LD2 and LD7 categories of language s take on a positively significant coefficient and with the inclusion of education interaction terms, we find negatively significant effects on income for the interaction terms for LD2, LD3 and LD7 languages, and positively significant effect on income for the interaction term for LD6 languages. These results highlight the differences in the effect of linguistic distance from English when interacted with countries' secondary school enrollment rates. More importantly, we can see that the effect of distance from English is not linear and that a larger distance from English, even after controlling for economic and political differences across countries, is not necessarily associated with lower levels of income. The results in column (2) with education interaction terms indicate that only the LD6 category of languages has a negatively significant effect on growth- however, once we control for the level of secondary school enrollment rates, the interaction coefficient for the LD6 category becomes positive. This finding suggests that we should consider total effects of linguistic distance measures to fully assess the relationship between linguistic distance categories and growth.

Columns (3) and (4) display the inclusion of trade openness interaction terms on income and growth respectively. Similar to the baseline case, we find LD2 and LD7 categories to be positively significant for income, and LD2, and LD7 categories to be negatively significant for growth. Additionally, the LD3 category of languages are now found to take on a significantly positive coefficient in both our income and growth regressions. This suggests that taking into account the level of trade openness and further examining the link between linguistic distance and income and growth may be unraveling further positive effects for countries belonging to this linguistic category. Interestingly, the trade openness dummy for the LD3 category is negatively significant for both income and growth, which calls for a deeper examination of total effects of linguistic distance categories.

**Table 5: Baseline with Interaction Terms**

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	Level	Growth	Level	Growth	Level	Growth
LD1	0.0495 (0.258)	-2.086 (1.533)	0.0648 (0.180)	-0.469 (0.690)	0.180 (0.264)	-1.910 (1.566)
LD2	1.714*** (0.565)	-5.393 (5.320)	0.494*** (0.188)	-3.209** (1.385)	1.497*** (0.311)	-10.38** (4.923)
LD3	0.581 (0.354)	-0.0724 (1.973)	0.537** (0.232)	2.771*** (0.860)	0.945** (0.390)	1.797 (2.100)
LD4	-0.0581 (0.347)	-0.957 (2.062)	0.134 (0.213)	-0.693 (0.838)	-0.0483 (0.339)	-0.901 (2.124)
LD5	0.152 (0.413)	-2.803 (1.752)	-0.00217 (0.372)	-2.189 (1.405)	0.169 (0.421)	-3.002 (2.074)
LD6	-0.241 (0.385)	-4.152** (1.870)	0.120 (0.353)	-1.973 (1.444)	-0.295 (0.432)	-4.974** (1.954)
LD7	8.931*** (2.042)	-27.35 (20.21)	0.696*** (0.159)	-3.089*** (0.897)	3.659*** (0.875)	7.659 (9.532)
SecEnroll	0.00236 (0.00194)	-0.00874 (0.0168)	0.00196** (0.000897)	0.00637 (0.0105)	0.00223 (0.00185)	-0.00862 (0.0174)
LD1 x SecEnroll	-0.00243 (0.00219)	0.0155 (0.0156)			-0.00166 (0.00207)	0.0179 (0.0167)
LD2 x SecEnroll	-0.0115** (0.00540)	0.0344 (0.0549)			-0.00963*** (0.00319)	0.0709 (0.0518)
LD3 x SecEnroll	-0.00562* (0.00340)	-0.0185 (0.0280)			-0.00614* (0.00314)	0.0135 (0.0243)
LD4 x SecEnroll	0.00268 (0.00346)	0.00248 (0.0222)			0.00249 (0.00355)	0.00155 (0.0256)
LD5 x SecEnroll	-0.00251 (0.00230)	0.0121 (0.0174)			-0.00257 (0.00210)	0.0101 (0.0186)
LD6 x SecEnroll	0.00637*** (0.00228)	0.0542** (0.0232)			0.00612*** (0.00226)	0.0423* (0.0223)
LD7 x SecEnroll	-0.0849*** (0.0201)	0.254 (0.200)			-0.0302*** (0.00833)	-0.108 (0.0897)
TradeOpen	0.00147*** (0.000470)	0.0126*** (0.00353)	0.00238*** (0.000614)	0.0161*** (0.00391)	0.00233*** (0.000587)	0.0164*** (0.00435)
LD1 x TradeOpen			-0.00261*** (0.000851)	-0.00392 (0.00614)	-0.00242*** (0.000804)	-0.00503 (0.00662)
LD2 x TradeOpen			0.000775 (0.00131)	0.0112 (0.0110)	0.000300 (0.000916)	0.0142 (0.0104)
LD3 x TradeOpen			-0.00370*** (0.000923)	-0.0391*** (0.0124)	-0.00380*** (0.000949)	-0.0401*** (0.0121)
LD4 x TradeOpen			0.000463 (0.00115)	-0.00353 (0.00658)	9.53e-05 (0.00121)	-0.00138 (0.00805)
LD5 x TradeOpen			-0.00101 (0.000939)	0.00114 (0.0117)	-0.000408 (0.000724)	0.00132 (0.0120)
LD6 x TradeOpen			0.00136 (0.00166)	0.0233 (0.0168)	0.000622 (0.00160)	0.0174 (0.0168)
LD7 x TradeOpen			-0.0102*** (0.00249)	0.0463* (0.0259)	-0.00776*** (0.00153)	0.0581*** (0.0167)
Constant	5.128*** (0.393)	10.42*** (2.000)	5.212*** (0.346)	9.632*** (1.652)	5.088*** (0.386)	10.70*** (2.117)
Observations	1,583	1,582	1,583	1,582	1,583	1,582
Number of Countries	97	97	97	97	97	97

Robust standard errors in parentheses

\*\*\* p&lt;0.01, \*\* p&lt;0.05, \* p&lt;0.1

Looking at the final two columns from Table 5, which include both the education and trade openness interaction terms, we find that the education interaction terms are negatively significant for the LD2, LD3 and LD7 categories and positively significant for the LD6 category on levels of income. This result shows that controlling for the gross enrollment in secondary education across countries, the distance of languages from English has a non-linear effect on levels of income. In terms of the impact of

these education interaction terms on growth, we find minimal significant effect, with only the LD6 category being positive and significant. The interaction terms of trade openness with linguistic distance display a negatively significant effect on income for LD1, LD3 and LD7 categories. Similarly, the LD3 category has a significantly negative impact on growth rates. This suggests that linguistic distance from English affects income and growth of incomes through the trade channel, but

as with education, the link between linguistic distance and trade openness is not linear and tends to be more negative for countries using LD3 languages. The examination of coefficients of our linguistic distance categories displays similar patterns, with the LD2 category having a positive effect on income and a negative effect on growth, and LD7 having a positive effect on income. Additionally, we find the LD3 category to be positively significant for income levels and the LD6 category to be negatively significant for growth rates. To further understand the relationship between linguistic distance from English and income levels and growth through the education and trade channels, we calculate both the partial and total effects for our interaction terms to clarify the connection between linguistic distances and income levels and growth.

3.2.1. Partial effects by linguistic distance category

To calculate the total effect of our education and trade openness interaction terms with the linguistic distance measures, we use partial derivatives of income and growth with respect to education (secondary school gross enrollment rate) and trade openness. Adding these interaction terms into our baseline models in Equations 1 and 2 generates the following regression models:

$$\begin{aligned}
 Y_{j,t} = & \alpha + \delta_1 LD1 + \delta_2 LD2 + \delta_3 LD3 + \delta_4 LD4 + \delta_5 LD5 + \delta_6 LD6 \\
 & + \delta_7 LD7 + \beta_0 SecEnroll_{j,t} + \beta_1 \{LD1_j \times SecEnroll_{j,t}\} + \\
 & \beta_2 \{LD2_j \times SecEnroll_{j,t}\} + \beta_3 \{LD3_j \times SecEnroll_{j,t}\} + \\
 & \beta_4 \{LD4_j \times SecEnroll_{j,t}\} + \beta_5 \{LD5_j \times SecEnroll_{j,t}\} + \\
 & \beta_6 \{LD6_j \times SecEnroll_{j,t}\} + \beta_7 \{LD7_j \times SecEnroll_{j,t}\} + \\
 & \gamma_0 TradeOpen_{j,t} + \gamma_1 \{LD1_j \times TradeOpen_{j,t}\} + \\
 & \gamma_2 \{LD2_j \times TradeOpen_{j,t}\} + \gamma_3 \{LD3_j \times TradeOpen_{j,t}\} + \\
 & \gamma_4 \{LD4_j \times TradeOpen_{j,t}\} + \gamma_5 \{LD5_j \times TradeOpen_{j,t}\} + \\
 & \gamma_6 \{LD6_j \times TradeOpen_{j,t}\} + \gamma_7 \{LD7_j \times TradeOpen_{j,t}\} + \\
 & \sum_k \psi_k X_{j,t}^k + u_j + \epsilon_{j,t}
 \end{aligned}
 \tag{3}$$

where  $Y_{j,t}$  represents national income levels,  $LogGDPpc_{j,t}$  or economic growth,  $GDPpcGrowth_{j,t}$ . We can then calculate the partial derivatives, which help to evaluate the total effect of both channels on income and growth across countries.

For the education channel, we have:

$$\begin{aligned}
 \frac{\partial Y_{j,t}}{\partial SecEnroll_{j,t}} = & \beta_0 + \beta_1 LD1_j + \beta_2 LD2_j + \beta_3 LD3_j \\
 & + \beta_4 LD4_j + \beta_5 LD5_j + \beta_6 LD6_j + \beta_7 LD7_j
 \end{aligned}
 \tag{4}$$

Similarly, for the trade channel, we have:

$$\begin{aligned}
 \frac{\partial Y_{j,t}}{\partial TradeOpen_{j,t}} = & \gamma_0 + \gamma_1 LD1_j + \gamma_2 LD2_j + \gamma_3 LD3_j \\
 & + \gamma_4 LD4_j + \gamma_5 LD5_j + \gamma_6 LD6_j + \gamma_7 LD7_j
 \end{aligned}
 \tag{5}$$

Following this approach and using the results from Table 5, we calculate the partial derivatives of income levels and growth rates with respect to education and trade openness for each of our seven

Figure 2: Total effects on income levels by linguistic distance categories

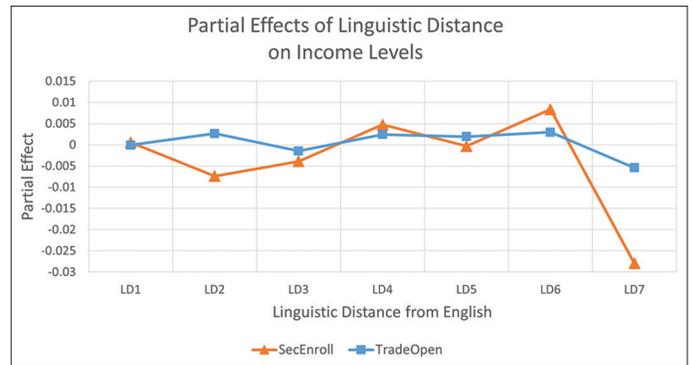


Figure 3: Partial effects on growth by linguistic distance categories

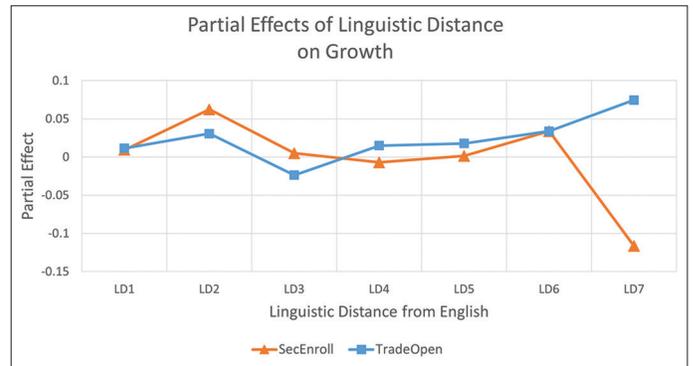


Table 6: Partial Effects: Linguistic Distance estimates

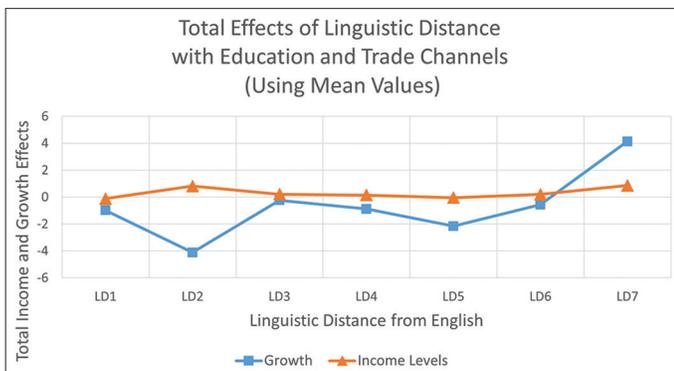
	$\frac{\partial LogGDPpc_{j,t}}{\partial SecEnroll_{j,t}}$	$\frac{\partial LogGDPpc_{j,t}}{\partial TradeOpen_{j,t}}$	$\frac{\partial GDPpcGrowth_{j,t}}{\partial SecEnroll_{j,t}}$	$\frac{\partial GDPpcGrowth_{j,t}}{\partial TradeOpen_{j,t}}$
LD1	0.00057	-0.00009	0.00928	0.01137
LD2	-0.00740	0.00263	0.06228	0.03060
LD3	-0.00391	-0.00147	0.00488	-0.02370
LD4	0.00472	0.00243	-0.00707	0.01502
LD5	-0.00034	0.00192	0.00148	0.01772
LD6	0.00835	0.00295	0.03368	0.03380
LD7	-0.02797	-0.00543	-0.11662	0.07450

linguistic distance categories. For example, the partial derivative of income levels with respect to education for the LD1 category is calculated to be  $\beta_1 + \beta_2 = 0.00057$ , while the partial derivative of GDP per capita growth with respect to education for LD1 category is 0.00928. Given the positive results of both partial derivatives, the LD1 category has an overall positive effect on income and growth via the education channel. Alternatively, the partial derivative of income levels with respect to trade openness for the LD1 category is calculated to be  $\gamma_1 + \gamma_2 = -0.00009$ , while the partial derivative of GDP growth per capita with respect to trade openness for LD1 category is 0.01137. These two partial effects suggest that the languages the least distant from English tend to have a negative effect through the trade channel on income levels; however, this effect becomes positive when examining growth rates. Table 6 reports these calculated partial effects for each of our seven linguistic distance categories. To visualize this relationship across linguistic distance categories, Figures 2 and 3 present the partial effect estimates for each of our seven LD variables. Overall, these figures demonstrate the expected negative effect of linguistic distance from English on income and growth rates with respect to the education channel. The partial effect for the LD7 category is found to take on the lowest value across all other linguistic distance categories under the education channel for both income levels and growth rates. However, when we consider the trade channel, the partial derivative for the LD7 category is found to be the highest for growth among all categories. This presents an interesting juxtaposition: even after controlling for differences in trade across countries, linguistic distance can still have positive effects beyond the explanation of economic control variables.

**Table 7: Total Marginal Effects: Linguistic Distance estimates**

	$\frac{\partial \text{LogGDPpc}_{j,t}}{\partial \text{LD}(i)_{j,t}}$	$\frac{\partial \text{GDPpcGrowth}_{j,t}}{\partial \text{LD}(i)_{j,t}}$
LD1	-0.124	-0.978
LD2	0.814	-4.116
LD3	0.209	-0.239
LD4	0.141	-0.892
LD5	-0.050	-2.163
LD6	0.200	-0.564
LD7	0.862	4.131

**Figure 4: Total effects by linguistic distance categories**



Delving deeper, several notable results emerge from the calculation of these partial derivatives across these two channels. First, for the education channel, the negative effect on income levels is most prominent for countries falling within the LD2, LD3, LD5 and LD7 linguistic groups, and the negative effect on growth is seen in the LD4 and LD7 categories. This reinforces that linguistic distance from English does not have a linear effect- countries using languages both closest to (LD1) and further from (LD6) English actually exhibit higher levels of income and growth, while most languages in the middle of the spectrum (but not LD4) exhibit lower levels.

Second, with respect the trade channel, the distance from English negatively influences income levels for countries in the LD1, LD3 and LD7 linguistic groups. When considering growth rates, we similarly find LD3 group to have a negative effect, while the remaining linguistic groups all display positive partial derivatives through the trade channel. The negative coefficients displayed by various linguistic groups suggest the existence of additional factors affecting countries' income levels within the trade channel. For example, regional trade agreements, or the relaxing of trade restrictions between countries could influence the direction in which languages affect trade and income levels across countries. A more detailed analysis could study the bilateral trade patterns across countries, taking into account the differing linguistic groups that countries fall under, but unfortunately, this aspect is beyond the scope of this paper.

**3.2.2. Total marginal effects by linguistic distance category**

Finally, we can also calculate the total effects of the linguistic distance measures, for each LD category, using the following partial derivatives, with the LD1 category given as an example:

$$\frac{\partial \text{LogGDPpc}_{j,t}}{\partial \text{LD1}_j} = \delta_1 + \beta_1 \text{SecEnroll}_{j,t} + \gamma_1 \text{TradeOpen}_{j,t} \quad (6)$$

$$= 0.180 + (-0.00166) \text{SecEnroll}_{j,t} + (-0.00242) \text{TradeOpen}_{j,t}$$

$$\frac{\partial \text{GDPpcGrowth}_{j,t}}{\partial \text{LD1}_j} = \delta_1 + \beta_1 \text{SecEnroll}_{j,t} + \gamma_1 \text{TradeOpen}_{j,t} \quad (7)$$

$$= -1.910 + (0.0179) \text{SecEnroll}_{j,t} + (-0.00503) \text{TradeOpen}_{j,t}$$

To evaluate the total effects for each of the seven linguistic distance categories, we use the mean values for *SecEnroll* and *TradeOpen*, as shown in Table 3. Table 7 reports these total effects for each of our seven linguistic distance categories.

To visualize the relationship of these total effects across linguistic distance categories, Figure 4 depicts the total marginal effects for each of our seven linguistic distance categories on both income levels and growth. We see that most of the linguistic distance categories have small positive effects on income levels, with the exceptions of LD1 and LD5 categories, which display small but negative effects on income levels across countries. Conversely, we find more variation across categories with respect to growth. Our results show that there is a negative relationship between linguistic distance from English and growth for the first six linguistic

**Table 8: Baseline with Interactions, Country Income Decomposition**

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Level	Growth	Level	Growth	Level	Growth	Level	Growth
LD1	0.742* (0.429)	-4.253 (2.712)	-1.201** (0.469)	-13.54* (7.630)	1.055*** (0.392)	-15.85*** (4.823)	0.678*** (0.214)	3.348 (2.422)
LD2							0.853*** (0.172)	-18.89** (7.852)
LD3	1.692*** (0.446)	17.59*** (3.951)			1.699*** (0.391)	-5.036 (5.610)		
LD4	1.183 (0.835)	-5.891 (3.667)	-0.890 (0.713)	-6.698 (7.929)	0.825 (0.717)	-11.98** (5.485)	0.629* (0.340)	10.51 (6.550)
LD5	1.226* (0.640)	-12.57 (7.894)	0.392 (0.440)	-5.689 (7.091)	-0.665 (0.664)	-13.85 (11.16)	0.271** (0.128)	-6.350*** (2.280)
LD6			-1.211*** (0.451)	-9.861 (7.284)	1.061* (0.600)	-30.18*** (8.457)	-1.288*** (0.290)	10.31 (11.76)
LD7							1.723* (0.887)	-3.593 (11.59)
SecEnroll	0.0132*** (0.00453)	-0.0200 (0.0608)	-0.00408 (0.00503)	-0.140 (0.0912)	0.0142*** (0.00346)	-0.133** (0.0631)	0.00186* (0.00102)	0.0237 (0.0161)
LD1 x SecEnroll	-0.00649 (0.00539)	-0.0683 (0.0580)	0.00924* (0.00542)	0.173* (0.0898)	-0.00829** (0.00398)	0.194** (0.0828)	-0.00388** (0.00169)	-0.0149 (0.0222)
LD2 x SecEnroll							-0.00583*** (0.00172)	0.157** (0.0774)
LD3 x SecEnroll I	-0.00971* (0.00536)	-0.0594 (0.0403)			-0.00703 (0.00565)	0.169* (0.0929)		
LD4 x SecEnroll	-0.00575 (0.00625)	-0.207** (0.0949)	0.0119 (0.00765)	0.0749 (0.0973)	-0.00748 (0.00662)	0.136** (0.0635)	-0.00498* (0.00273)	-0.0559 (0.0602)
LD5 x SecEnroll	-0.0117* (0.00630)	-0.0290 (0.0941)	0.00581 (0.00490)	0.0976 (0.0906)	0.00868 (0.00805)	0.202 (0.159)	-0.00248** (0.000987)	-0.0102 (0.0183)
LD6 x SecEnroll			0.0158*** (0.00526)	0.110 (0.0929)	-0.0156*** (0.00567)	0.327*** (0.109)	0.00711** (0.00345)	-0.0549 (0.136)
LD7 x SecEnroll							-0.0134 (0.00868)	0.00141 (0.109)
TradeOpen	0.00486** (0.00198)	-0.0416* (0.0247)	-0.000734 (0.00263)	0.0176 (0.0155)	0.00316** (0.00123)	0.0291* (0.0170)	0.00170** (0.000699)	0.0336*** (0.00488)
LD1 x TradeOpen	-0.00628*** (0.00220)	0.0583 (0.0372)	0.00166 (0.00287)	0.0162 (0.0215)	-0.00363 (0.00288)	-0.0103 (0.0364)	-0.00129* (0.000705)	-0.0302*** (0.00607)
LD2 x TradeOpen							0.000328 (0.000623)	0.00671 (0.00980)
LD3 x TradeOpen	-0.00872*** (0.00229)	-0.358*** (0.0543)			-0.00689** (0.00278)	-0.0834*** (0.0289)		
LD4 x TradeOpen	-0.00750 (0.00555)	0.213*** (0.0723)	-7.14e-05 (0.00409)	0.00976 (0.0255)	0.000973 (0.00203)	0.00578 (0.0197)	-0.00245 (0.00160)	-0.0470*** (0.0122)
LD5 x TradeOpen	-0.00443** (0.00221)	0.107*** (0.0187)	-0.000887 (0.00292)	-0.0183 (0.0197)	-0.00174 (0.00145)	-0.0169 (0.0345)	0.00246*** (0.000815)	0.0835*** (0.0101)
LD6 x TradeOpen			-0.00353 (0.00329)	0.0455 (0.0281)	-3.95e-05 (0.00166)	0.0522* (0.0287)	0.00126 (0.00165)	-0.0544* (0.0285)
LD7 x TradeOpen							-0.00470** (0.00191)	0.0757*** (0.0204)
Constant	4.524*** (0.619)	16.44*** (4.731)	6.347*** (0.522)	16.98** (7.607)	6.197*** (0.808)	41.62*** (10.84)	4.108*** (0.617)	-11.17 (8.614)
Observations	203	203	404	404	348	347	628	628
Number of Countries	22	22	26	26	18	18	31	31

Robust standard errors in parentheses

\*\*\* p&lt;0.01, \*\* p&lt;0.05, \* p&lt;0.1

distance categories (with LD2=German being an even larger negative outlier). However, this relationship reverses, with LD7 displaying a positive total effect on growth (with LD7=Japanese being a relatively large positive outlier).

Overall, our results suggest that the relationship between our linguistic distance measures and our dependent variables is nonlinear. Additionally, the non-linearity is also discontinuous,

displaying abrupt changes for total effects across categories of linguistic distance on growth rates. In terms of total effects, we find the LD2 category, which consists of countries using the German language to incur the greatest negative effect of linguistic distance from English on GDP per capita growth rate. In contrast, the most linguistically distant language category, LD7 which consists solely of Japanese, is found to have a positive link to growth. These two results allude to the possibility that countries with high income

levels may have an additional influence which may not be captured by our current regressions. To account for this possibility, we consider sub-sampling our data by income groups.

### 3.3. Baseline by National Income Groupings

While our analysis allows for a robust selection of countries in our dataset, growth theory suggests there may be differential effects of these control variables on average incomes and growth rates depending on countries' current relative national income levels. We therefore decompose our analysis across income groupings- Low Income, Lower-Middle Income, Upper-Middle Income, and High Income countries- as categorized by the World Bank's country classifications.<sup>11</sup> This decomposition breaks our 97 countries into four roughly equally-sized groupings of between 18 and 31 countries. Unfortunately, this decomposition results in some of our linguistic distance from English metrics being omitted from various regressions specifications, due to lack of observations. For example, this is most easily observed in the LD2 and LD7 categories, representing only German- and Japanese-speaking countries respectively, which are all found in the High Income grouping. Similarly, with only 3 languages in the category, there are no LD3 category countries in the Lower-Middle or High Income groupings. Perhaps more surprisingly, despite encompassing 17 different languages, there are no LD6 countries in the Low Income grouping. Table 8 presents the results of this decomposition.

Examining the results for Low Income countries, we observe that three of the four applicable linguistic distance categories in this grouping, LD1, LD3 and LD5 are positive and significant for our income level regression in column (1). This suggests that in this grouping, most countries with a medium of instruction other than English actually exhibit higher mean income levels than those that predominantly use the English language. Looking at the growth regressions in column (2), we find that the effect of linguistic distance changes, with the closest (LD1) and most distant (LD5) languages having a predicted negative effect, although these coefficients are not statistically significant. However, we also observe that LD3 category languages tend to bring positive growth benefits to Low Income countries. Columns (3) and (4) show that LD1 category languages have negatively significant effects on both income levels and growth for Lower-Middle Income countries. The LD6 category is also shown to have a negative impact on income levels in this income grouping. Similar to the Low Income countries, these findings suggest that the closest (LD1) and the most distant (LD6) languages to English tend to have the most statistically significant effects on income levels in Lower-Middle Income countries.

The results from columns (5) and (6) for Upper-Middle Income countries demonstrate that the closest linguistically distant languages in the LD1 category have a positive effect on income levels and a negative effect on growth rates. Similarly, the LD3 and LD6 categories are found to positively influence income levels in Upper-Middle Income countries, whereas the LD4 and LD6 categories are both shown to have a negatively significant effect on growth rates. These results from Upper-Middle Income

countries are more supportive of the notion that further distance from English can create disadvantages in growth rates. This is reinforced to some degree by the High Income country results in columns (7) and (8), which emphasizes once again the non-linearity of the impact of linguistic distance measures on growth, with the effect changing from negatively significant to positively significant as we move from LD2 to LD4 category.

Overall, the results from decomposing the countries according to their income levels demonstrate that the linguistic distance variables do not have a linear impact on the level of income or its growth rate. An examination of linguistic distance categories reveals that proximity of languages to English tend to have more positive effects on income and in some cases on growth for Upper-Middle and High Income countries. The negative effects of linguistic distance from English are shown to be the largest on growth for Upper-Middle Income countries, which are found to have the most statistically significant results.

## 4. CONCLUDING REMARKS

English has increasingly become a language of instruction across a wide variety of countries. It has been the most prominent language of communication around the world. The commonality of the use of English has been reported to lower transaction costs and generate positive externalities. With growing linkages across countries through trade and globalization, the English language could be considered a possible factor that affects education, trade, income, and economic growth. To examine if this has been the case, in this paper we inspect whether the distance from English language has affected income levels and economic growth across countries and over time, as demand continues to grow for the English language in both educational and professional settings on the international stage.

Our baseline specification, using seven categories of linguistic distance from English (LDE) measures finds the LD2 (German) and LD7 (Japanese) categories to have a positively significant effect on income levels but a negatively significant effect on growth rates, while most other linguistic distance categories do not appear to have a statistically significant effect on incomes and growth. To investigate whether this may be obfuscated by the use of multiple languages of instruction in some countries, we include a dummy variable to account for multiple languages, but we do not find any statistically significant changes in the results from our benchmark model. Adding regional dummy variables to control for differences across countries similarly does not noticeably alter our previous findings.

In understanding the impact of linguistic distances from English, we consider the two specific channels through which languages may affect income and growth rates across countries. We include education and trade interaction terms, and our results suggest that the LD2 and LD7 categories have a positively significant impact on income levels; however, this becomes negatively significant for both categories when examining economic growth. Under the education channel, further distance does not imply lower income. The case of LD6 languages demonstrates that further distance can in fact lead to a positive effect on incomes and growth. This implies that there is a non-linearity and discontinuity (across different

<sup>11</sup> Available at <https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups>

language categories and how they affect incomes and growth) in the education channel. The cost of acquisition, as a result, does not necessarily suggest that more distant languages from English will bring about lower income levels. With the trade channel, we find more significant effects of linguistic distance from English. With the inclusion of trade interaction terms, we can no longer argue that further distance from English leads to lower levels of income or growth. Additionally, we find that these linguistic distance measures have highly non-linear and discontinuous effects on incomes and growth, regardless of the channels examined. The impact of distance from English does not increase linearly across categories of languages, which are classified as being more distant from the English language. Our interaction terms display an interesting relationship, where the linguistic distance measures are more often found to be more significant through the trade channel than the education channel. Furthermore, the interaction terms for education are mostly shown to have a negative effect on income levels but not on growth, whereas the interaction terms for trade are found to have similar impacts on both income levels and growth rates. A more detailed analysis shows that the partial and total effects of linguistic distance measures tell a similar story.

Examining the partial effects for education, as expected, we find that the languages that are the least distant from English have a positive effect, and the languages most distant from English have a negative effect, on both income levels and growth rates. Conversely, the partial effects for trade for the LD1 category are shown to have a negative impact on income levels and a positive impact on growth rates. The education channel suggests lower levels of income and growth in language categories further from English. However, the increasing linguistic distance, once again, does not follow a linear pattern, with some more distant categories reporting positive partial effects under the education channel for both income levels and growth while others do not. The most distant language category, LD7, on the other hand, is found to have the largest negative partial effects on both income levels and growth under the education channel. With the trade channel, the most distant language category is found to have a positive impact on growth suggesting that the most linguistically distant language from English can still have positive effects beyond the effects of our control variables.

The total effects of linguistic distance categories on income levels and growth rates demonstrate small but positive effects of five categories of linguistic distance on income levels. For the growth rates, all but one of the categories are shown to have a negative impact. Once again, we find evidence for a non-linear relationship between linguistic distance and growth, with the effect being the largest for the LD2 category. Conversely, the LD7 category (consisting of Japanese) is shown to have a positive effect on growth rates. These interesting results imply that country specific effects, which may not have been captured by our control variables, could be influencing our baseline regressions. To account for this possibility, we sub-sample our dataset by income groups. Our results show that languages closer to English have the largest positive effects on income levels in Low Income, Upper-Middle-Income and High-Income countries. The largest negative effects on growth are found to be in Upper-Middle-Income countries.

Overall, our results demonstrate that linguistic distance affects income levels and growth non-linearly across countries. Further distance from English does not necessarily translate into declining levels of income or growth rates in our dataset. Conversely, we find that, through the trade channel, more distant languages tend to experience positive rates of growth. This brings forth the possibility of linguistic distance being influenced by other factors, such as bilateral or multilateral trade agreements across countries.

Our study is the first to examine the channels through which linguistic distance from English can influence income and growth across countries and over time. We provide a more detailed approach in constructing linguistic distance categories from English. Future work could further examine other channels through which linguistic distance may affect growth. As more data on schooling and educational attainment become available, particularly for developing countries, we can analyze whether school attendance or completion could be a stronger indicator for the education channel. The availability of more cultural variables across countries could help control for some additional country-specific effects that may be contributing to the non-linearity of our results.

## REFERENCES

- Adsera, A., Pytlikova, M. (2015), The role of language in shaping international migration. *The Economic Journal*, 125(586), 49-81.
- Chiswick, B.R., Miller, P.W. (2005), Linguistic distance: A quantitative measure of the distance between English and other languages. *Journal of Multilingual and Multicultural Development*, 26(1), 1-11.
- Chiswick, B.R., Miller, P.W. (2012), Negative and positive assimilation, skill transferability, and linguistic distance. *Journal of Human Capital*, 6(1), 35-55.
- Chiswick, B.R., Miller, P.W., editors. (2015), International migration and the economics of language. In: *Handbook on the Economics of International Migration*. Vol. 1A. Amsterdam, Netherlands: Elsevier. p211-269.
- Desmet, K, Ortuno-Ortin, I., Weber, S. (2009), Linguistic diversity and redistribution. *Journal of the European Economic Association*, 7(6), 1291-1318.
- Education First. (2020), EF English Proficiency Index: A Ranking of 100 Countries and Regions by English Skills. Available from: <https://www.ef.com/assetscdn/WIBIwq6RdJvcD9bc8RMD/cefcom-epi-site/reports/2020/ef-epi-2020-english.pdf> [Last accessed on 2021 Sep 29].
- Foreign Service Institute (FSI). (2020), Foreign Language Training. Available from: <https://www.state.gov/foreign-language-training> [Last accessed on 2021 Sep 29].
- Goren, E. (2018), Consequences of linguistic distance for economic growth. *Oxford Bulletin of Economics and Statistics*, 80(3), 625-638.
- Hart-Gonzalez, L., Lindemann, S. (1993), Expected Achievement in Speaking Proficiency. Mimeo: School of Language Studies, Foreign Services Institute, Department of State, Mimeo.
- Hutchinson, W.K. (2005), Linguistic distance as a determinant of bilateral trade. *Southern Economic Journal*, 72(1), 1-15.
- Ispording, I.E. (2014), Disadvantages of linguistic origin evidence from immigrant literacy scores. *Economics Letters*, 123(2), 236-239.
- Ispording, I.E., Otten, S. (2013), The costs of Babylon linguistic distance in applied economics. *Review of International Economics*, 21(2), 354-369.

- Ku, H., Zussman, A. (2010), *Lingua franca: The role of English in international trade*. *Journal of Economic Behavior and Organization*, 75(2), 250-260.
- Ly, A., Esperança, J., Davcik, N.S. (2018), *What drives foreign direct investment: The role of language, geographical distance, information flows and technological similarity*. *Journal of Business Research*, 88, 111-122.
- Marsh, D. (2006), *English as a Medium of Instruction in the New Global Linguistic Order: Global Characteristics, Local Consequences*. Jyväskylä, Finland: UNICOM, Continuing Education Centre, University of Jyväskylä, Finland. Available from: <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.125.2388&rep=rep1&type=pdf>
- Melitz, J. (2008), *Language and foreign trade*. *European Economic Review*, 52(4), 667-699.
- Ndamba, G.T., Van Wyk, M.M., and Sithole, J.C. (2017), *Competing purposes: Mother tongue education benefits versus economic interests in rural Zimbabwe*. *International Indigenous Policy Journal*, 8(1), 1-21.
- Rose, H., McKinley, J. (2016), *Realities of doing research in applied linguistics*. In: *Doing Research in Applied Linguistics*. Abingdon, United Kingdom: Routledge.
- Rubagumya, C.M. (1989), *English-medium instruction in Tanzanian secondary schools: A conflict of aspirations and achievements*. *Language, Culture and Curriculum*, 2(2), 107-115.
- Teorell, J., Sundstrom, A., Holmberg, S., Rothstein, B., Pachon, N.A., Dalli, C.M. (2021), *The Quality of Government Standard Dataset*. University of Gothenburg: The Quality of Government Institute. Available from: <https://www.gu.se/en/quality-government/qog-data> [Last accessed on 2021 Jun 17].
- The World Bank. (2021), *World Bank Development Indicators*. Available from: <https://databank.worldbank.org/source/world-development-indicators> [Last accessed on 2021 Sep 29].