LINGUISTIC STANDARDIZATION AND ECONOMIC GROWTH

Amy H. Liu
Assistant Professor
University of Colorado Boulder
amy.liu@colorado.edu

Elise Pizzi
PhD Candidate
University of Colorado Boulder
elise.pizzi@colorado.edu

Abstract: Do heterogeneous countries that have standardized their linguistic repertoires enjoy the same economic benefits more commonly found in their homogeneous counterparts? The advantages of ethno-linguistic homogeneity on economic growth is well documented. Intra-group interactions are more likely to be efficient because co-ethnics can gauge each other more accurately, interact more frequently, and can find each other more easily. We argue these purported advantages are not limited to those from the same ethno-linguistic group; instead, they can be enjoyed by those who have learned the language as a second language as well. Moreover, when this language corresponds to the official language of the state, this efficiency can parlay into higher growth levels. We examine this argument with two different tests: (1) a comparative case study of Indonesia and the Philippines; and (2) a statistical analysis using newly-compiled historical data from Ethnologue. The results suggest that a country with exogenously high levels of heterogeneity can avoid what Easterly and Levine (1997) call “the growth tragedy” by endogenously standardizing its linguistic repertoire.

July 1, 2014

Paper prepared for the International Political Science Association’s 2014 World Congress (Montreal, Canada). Please do not cite without permission.
Heterogeneity impedes growth is “one of the most powerful hypotheses in political economy.” Banerjee et al. 2005: 639; cited in Habyarimana et al. 2007: 709

1. INTRODUCTION

The economic effects of ethno-linguistic heterogeneity are well documented (Alesina et al. 2003; Easterly and Levine 1997; Posner 2004). Previous research suggests that because language demarcates ethnic group boundaries (Brown and Ganguly 2003), the presence of ethno-linguistic differences also reveal an inherent tension. In ethno-linguistically heterogeneous countries, there is diversity of preferences over whose welfare to consider, which public goods to produce, and to what extent to offer said goods. This conflict over the distribution of scarce resources results in under-spending of core public goods and over-spending of targeted private goods (Alesina et al. 1999; Alesina and LaFerrara 2005; Treisman 1999). Over time, such arrangements can be detrimental to any country’s economy. The robustness of this link between high levels of ethno-linguistic fractionalization and poor economic performance is evident in the opening epigraph.

While ethno-linguistic heterogeneity is in and of itself important, these analyses, by focusing strictly on language as ethnic marker, overlook the fact that language serves another purpose: It is a vehicle of communication. When two individuals speak the same language—whether that be any language—they are able to interact with relative efficiency. And since individuals can speak multiple languages (Laitin 1988; Pool 1992), there is no reason to believe the boundaries of a spoken language are strictly congruent with the boundaries of the ethnic group. Given that languages can transcend ethnic groups, a population at-large can interact in the same language, regardless of what other languages an individual can also speak, and thus economic exchanges flow with greater efficiency. In this paper, we examine this puzzle in detail: Do heterogeneous countries that have standardized their linguistic repertoires enjoy the same economic benefits more commonly found in their homogeneous counterparts?

By focusing on who can speak a language as opposed to who identifies culturally with the language, this paper calls attention to the efficiency-enhancing function of languages. We argue that when individuals—regardless of their ethnic identities—interact in the same language as the state, economic exchanges are more efficient. This efficiency, in turn, is conducive to economic growth. This is true even if the individual’s mother tongue (L1) is not that of the official state language. Instead, what matters is that the individual has had the opportunity to learn the official state language as a second language (L2).

This paper proceeds as follows. We begin by reviewing the literature on the economic effects of ethno-linguistic heterogeneity, highlighting the limitations of focusing strictly on language as an ethnic marker. We then introduce our argument and explain the logic underlying it: Language, as a vehicle of communication, can facilitate communication across ethnic groups. When the official state language is taught in schools and learned—even if only in the capacity as a second language—it can have the same effect on efficiency as ethno-linguistic homogeneity. We examine this argument first by looking at a comparative case study of Indonesia and the Philippines—two highly heterogeneous countries that started similarly but since then have diverged quite substantially in their economic performance. We suggest this discrepancy is in part due to differences in the degree of linguistic standardization. To further assess this claim, we use original data to statistically test whether the size of the population that speaks the official language has an effect on economic growth. This newly-constructed cross-sectional over-time measure takes into consideration not only those whose mother tongue is the official language but also those who have learned it as a second language. The results indicate linguistic
standardization has an important effect on economic growth. We conclude in the final section by noting that while homogeneity may have its inherent exogenous advantages, countries with high heterogeneity levels do have alternative endogenous channels to avoid the poverty trap.

2. ETHNIC HETEROGENCEITY AND ECONOMIC GROWTH
The link between ethno-linguistic heterogeneity and the “growth tragedy” (Easterly and Levine 1997: 1203) is well established. Easterly and Levine (1997) find that across Africa GDP growth rate is negatively associated with ethnic diversity. Particularly, they suggest ethno-linguistic heterogeneity obstructs the provision of public goods. Other scholars have drawn similar conclusions not simply about economic growth (Alesina et al. 2003; Posner 2004), but also about public goods provision (Alesina and La Ferrara 2005; Habyarimana et al 2007; Miguel and Gugerty 2005). Non-excludable public goods are likely to be lower as heterogeneity increases even though overall public spending may not necessarily be lower (Kimenyi 2006). This finding is by no means an Africa story. Even in OECD countries, which are on average much wealthier than their African counterparts, ethno-linguistic diversity is a significant—but negative—predictor of economic performance (Patsiurko et al. 2012). The strength of this association between ethnic heterogeneity and low economic growth is so robust that Posner (2004) notes, “[t]hanks largely to their article [Easterly and Levine (1997)], it is now de rigueur for economists to include a measure of ethnic diversity in their cross-country growth regressions” (849). Other examples include Mauro (1995) and Przeworski et al. (2000).

There are multiple explanations why ethnic heterogeneity stunts economic growth. First, ethnic groups may have different preferences and may not agree over policy priorities. For example, an ethnic group may oppose public education if it is perceived to benefit another ethnic group more (Alesina et al. 1999). Interestingly, using experimental data, Habyarimana et al. (2007; 2009) find no evidence for divergent preferences among ethnic groups. Instead, they find tension over the specific type or location of public services the community or government should provide. For example, an ethnic group may support the construction of hospitals and other public service facilities but only when they are located in places that benefit the ethnic group of concern (Kimenyi 2006).

Another possibility is that individuals behave differently depending on with whom they are interacting. That is, their behavior may be more cooperative and they may be more willing to collaborate with members of their own ethnic group. This may be related to a lack of interpersonal trust across ethnic groups (Alesina and La Ferrara 2005) or doubts about the ability to enforce cooperative agreements across ethnic groups (Miguel and Gugerty 2005). When members of the in-group identify threats—whether actual or perceived—from the out-group, they assign the out-group to negative classifications (Fearon and Laitin 1996; Tajfel 1982).

Alternatively, cooperation among people from different ethnic groups may be more difficult because they do not share norms of collaboration, understandings, or rules that can guide interactions (Habyarimana et al. 2007, 2009). The idea that diverse communities lack some technology of cooperation is especially relevant when we look at linguistic diversity. When individuals cannot directly communicate with each other because of a language barrier, interactions are likely to be more challenging. Similarly, linguistic minorities that do not speak the official language of the state may find interactions with government institutions especially problematic (e.g., Liu and Baird 2012).

There are at least two limitations with this extant literature. The first has to do with conceptualization. The predominant focus on language is that of an ethnic marker. While a
language can be an ethnic marker, it is also a tool for communication and nation building (Laitin 1988). Knowing a second or third language can facilitate efficiency in communication and allow for cooperation regardless of identity. Moreover, being able to speak the official language of the state affords individuals opportunities. To put it differently, regardless of an individual’s racial or religious identity, the transaction costs of not knowing the language of the state are high. This is true even if the individual can speak the language of the majority. Consider Haiti. Although the Haitian population is linguistically homogeneous, the spoken language (Haitian Creole) is very distinct from the official language of the state (French). This disparity was evident in the aftermath of the 2010 earthquake when French-speaking aid workers arrived in Port-au-Prince but could not communicate effectively with the population at-large. Language learning is additive; knowing a second language does not imply that individuals are changing their identities or that there is “social consensus or harmony” (Laitin 1988: 293) but it does lead to more efficient interactions. We argue this efficiency is important in facilitating economic growth.

The second limitation with the extant literature has to do with measurement. Because the literature conceptualizes language as an ethnic marker, the boundaries of the language must by definition match that of the ethnic group (i.e., those whose mother tongue is strictly that language). For example, one of the most frequently used measures of ethnic heterogeneity is the Ethno-Linguistic Fractionalization (ELF) index (see Easterly and Levine 1997). The ELF uses a Herfindahl concentration index to measure the likelihood that two people, selected at random, will be from two different ethnic groups. Data for this measure come from the Soviet 1964 Atlas Narodov Mira. Although the ELF index is commonly employed, there are a number of concerns with its use. Posner (2004) identifies three limitations with this measure. First, it confounds the different effects of each ethnic marker (Alesina et al. 2003); second, it draws on outdated Soviet data (Campos and Kuzeyev 2007); and third, it measures the presence of any ethnic group based on demographics and not its political relevance (Gurr 1993).

Here, we identify a fourth limitation with the ELF. By focusing on language as a demographic ethnic marker, the ELF treats the number of speakers of any language as static over time. Language, however, is dynamic (Cederman, Wimmer, and Min 2010). Not all individuals are monolingual. Moreover, second (or third) languages can be learned. As more people study a language, the number of speakers increases as well—even if demographically, the group whose claim to the language as a mother tongue does not change. In this paper, when measuring the number of speakers for a language, we focus not only on group identity but also take into consideration the population of second language (L2) learners.

3. LINGUISTIC STANDARDIZATION AND EFFICIENCY

All governments rely heavily on their education system to advance their language planning visions (Albaugh 2014; Fishman 1989), save a few rare exceptions. In some instances, this reliance is absolute (Kaplan and Baldauf 2003). Regardless of language regime type (Liu 2014), at least one official language is studied by some qualified supermajority of the student-age population. Consider a power-concentrating language regime such as Thailand’s where the language of the King and the royal family (“central Thai”) is the only one recognized by the state. The fact that central Thai was by law renamed “standard Thai” is evidence of the political dominance of the one language. From a normative language justice standpoint, speakers of minority languages are unfairly denied the opportunity to learn their culture in schools. With

1 One such exception is Cambodia 1975-1979. The Khmer Rouge, attempting to return the country back to a classless agrarian state, closed down schools, burned books, and killed teachers (Short 2006).
respect to efficiency, however, they are learning the language of government and commerce. So while the demographic size of “central Thai” is a smaller percentage, the proficient population (either as a first or second language) is much larger.

Consider another language regime: Switzerland’s. In Switzerland, linguistic power is shared across multiple languages: German (63.7%), French (20.4%), Italian (6.5%), and Romansch (0.5%). While each canton has the right to dictate which (set) of the four languages is official at the sub-national level, the national Ministry of Education requires all Swiss students to learn two other languages. For almost the entire non-German native population, German is one of the two chosen languages (Education, Audiovisual and Cultural Executive Agency 2008). If we combine the native German speakers with the second language speakers, this would produce a supermajority of the population.

There is a third language regime: power-neutralization. These regimes are characterized by the recognition of a lingua franca—a language of inter-ethnic communication. The use of such languages by definition cuts across all ethno-linguistic group boundaries. In Indonesia, for instance, the official language is Indonesian, a language whose linguistic origin traces back to Malay. Malay, the mother tongue of a very small subset of the population, was widely spoken throughout the archipelago because of commercial activities dating back to Dutch colonial days (Bertrand 2003; Montolalu and Suryadinata 2007).

Language regime type notwithstanding, when parts of a population cannot interact with each other—and more importantly with the government—using the same language, the volume of necessary translations increases and the price for these translations becomes costly (Ginsburgh and Weber 2011). Failure to pay and reduce these costs can be economically detrimental on the long term. When each market transaction requires all relevant actors to re-specify similar conditions of previous transactions, this is highly inefficient. In contrast, when individuals can interact with each other and with the government in the same language, the number of translations necessary is rendered minimal. Given this discussion, we hypothesize the following:

**Hypothesis 1:** Countries with larger population of official language speakers have higher growth rates.

There are two different ways for a country to have a sizable population of official language speakers. The first is that the country is largely homogeneous and the language most people speak at home happen to be the language of the state. South Korea is one example, where over 99% of the population claim Korean—the official language of the state—as their mother tongue. The other way for a country to have a population proficient in the state language is to teach it in schools. If education is the apparatus for broadcasting the government’s vision of how the state should look, including the language of public service (Althusser 1971), it is essential that there are consumers of this ideology—i.e. students.

The link between education and economic growth is well established. Education is an important arena for human capital accumulation (Baum and Lake 2003; Brown and Hunter 1999). Directly, education increases human capital by teaching skills that enhance productivity (Gradstein, Justman, and Meier 2005). And indirectly, education reduces the inequality of pay (Birdsall, Ross, and Sabot 1997). It is no coincidence that education was the second largest expenditure (after defense) in Singapore, South Korea, and Taiwan—three of the economic “tigers” noted in the 1993 World Bank “East Asian Miracle Report.” Literacy is clearly important. In fact, no country has developed without a literate workforce (Easterly 2003). In this paper, we go one step further and call attention to the fact that (1) literacy in the official state
language is one such skill for enhancing productivity and (2) literacy in any language is not sufficient but rather literacy in the same language is necessary.

When a country has a large population—this includes boys and girls—attending school for a substantive number of years (i.e., through secondary education), it produces a significant population that has been exposed to the official language either as a first language (L1: instruction in official language) or second language (L2: instruction of official language). So while the demographic population of people who can claim L1 as a mother tongue is largely fixed, the country can increase this population through the numbers in L2. When the total L1 plus L2 population is substantive, this suggests there is one language that cuts across the population at large.

It is important, however, to identify whether this language corresponds with the language of the state. When the two languages do not match, those who are not literate in the official language are considered disenfranchised (Ginsburgh and Weber 2011). They are not able to enter into contractual agreements fairly. They are also not able to use the judicial system meaningfully to enforce violations of these agreements. And they are not able to understand the laws pertaining to taxes. In sum, it is not sufficient that the population is demographically homogeneous. Consider the previous example of Haiti. The language commonly spoken by the population is Creole, which—as many aid workers found out in the aftermath of the 2010 earthquake—is mutually unintelligible from French. The population, while homogenous, had very little exposure to the French language in the classrooms. This discussion leads us to the following hypothesis:

**Hypothesis 2:** Heterogeneous countries with larger student-age populations learning the official language of the state have growth rates no different from those of homogeneous countries.

## 4. INDONESIA VERSUS THE PHILIPPINES: COMPARATIVE CASE STUDIES

A comparison of Indonesia and the Philippines—two non-British colonies in Southeast Asia—can help elucidate how the presence of a standardized linguistic repertoire can facilitate efficiency. From a theoretical standpoint, these two cases lend themselves naturally to a most similar design (George and Bennett; Seawright and Gerring 2008). For instance, as noted in Table 1, neither country existed prior to the arrival of the Europeans (the Dutch in Indonesia and the Spaniards in the Philippines). Then, both countries were merely a scattered collection of islands across a broad, ethnically diverse archipelago. Upon independence, this lack of a shared collective identity proved to be a challenge in the early days of state-building. These struggles were magnified in light of their struggling economies after World War 2: The two countries found themselves with empty coffers and torched farmlands—no trivial matter given their dependence on agriculture. At the time, neither county could sustain its own rice production. While both countries flirted with democratic structures in the first decades, the 1960s would usher in a long-ruling autocrat in both (Suharto in Indonesia and Ferdinand Marcos in the Philippines). It is also fitting to note both men were anti-communists in their political ideology.

Yet despite their similar paths dating back to colonialism, their economic trajectories could not have been any more different—at least up until 1998 (see Figure 1). Under Suharto, Indonesia became a net exporter of rice and its economic growth would garner the attention of the World Bank as a “high performing Asian economy” (HPAE). In contrast, the Philippines remained “the sick man of Asia”. In fact, in that same report that designated Indonesia as an
HPAE, the Philippines was the only capitalist (i.e., non-communist) Asian country omitted from the World Bank report. This divergence is even more shocking considering that the Philippines had an additional factor that should have favored its economic performance: the American legacy. When the US took control of the Philippines after the Spanish-American War, this gave the Philippines at least two advantages over its Indonesian counterpart. First, the US had offered generous loans and other economic assistance to the Philippines. This should have helped with infrastructural development and other post-war reconstruction efforts. Second, despite President McKinley’s best efforts to recognize the linguistic diversity of the Philippines, the common language from the courtrooms to the classrooms during the US occupation was English. Consequently, the Philippines inherited a working language that was simplistic in its orthography, developed in its vocabulary, and global in its scope. This should have helped reduced transaction costs and facilitated economic exchanges.

[Figure 1 about here]

Yet despite this significant advantage, the Philippines chose not to standardize its linguistic repertoire in English but in Tagalog, the language of the plurality. The chauvinism displayed by the Tagalog-speakers to push through their language as the working language of the state drew extreme hostility from speakers of the other languages, most notably the Cebuano Bisayans. The opposition continued despite government efforts to rename the language “Pilipino” and then later “Filipino” in hopes of removing any ethnic overtones (Gonzalez 1999). Despite government desires for linguistic standardization in Tagalog, the reality was that the language used in the classrooms did not reflect the language envisioned by the government. Instead, outside of Manila, the vernacular languages were being used extensively without any credit from the government (Gonzalez 2003). This effectively further fractionalized the different ethnic groups.

In 1973 a new constitution was adopted whereby the education curriculum would now operate in bilingualism whereby one of the languages would be in English. This signaled a shift towards standardizing the linguistic repertoire in English. However, with dropout rates as high as 50% at the primary level in the non-Tagalog speaking areas (Gonzalez 1981), standardization efforts still did not materialize. Students were not being exposed sufficiently to the government-designated language. Instead, it only furthered the confusion in social, economic, and political settings (Roberts and Kitao 1987).

In contrast, the linguistic arrangement in Indonesia was one of great success. Despite there also being a language (Javanese) spoken by a sizable population around the capital, the government opted to employ another language in building the Indonesian state. Malay was renamed bahasa Indonesia to give it a national feel. Here, it is important to acknowledge that while Malay had been widely spoken throughout the region because of trade, the derivative spoken across Indonesia (bazar Malay) was quite different from the one that ended up being named Indonesian. So while one can argue it was a logical choice, it was neither a given outcome nor a choice easy to implement. The language needed to be standardized, and moreover, since it lacked a depth in vocabulary, many loan words needed. In short, the process was arguably more challenging than that of English in the Philippines. For a number of reasons, both political and practical, the Javanese in Indonesia did not insist on having their language be designated the national/official language of the state. In doing so, this allowed the Indonesian government to standardize its linguistic repertoire through Indonesian. Upon assuming office, Suharto designated Indonesian the exclusive medium of instruction in schools. To their detriment, the other vernacular languages were all banned from the classrooms. However, unlike
the Philippines, the language taught in the classrooms mirrored the language designated by the government. Moreover, enrollment rates were high (see World Bank 1993)—thereby exposing a generation of students to the standardized language (Bertrand 2003).

It is hard to overlook how differences in linguistic standardization have mattered for economic growth in the two countries today. Admittedly, a disclaimer warrants discussion. A number of factors can explain a country’s economic performance. Whether a language is able to facilitate efficiency and reduce transaction costs is most certainly not the only variable of importance. Our argument that linguistic efficiency played an important role in Indonesia’s growth does not by any means take away from Suharto’s other macro-economic policies. For instance, upon taking office after the 1965 coup, Suharto reversed many of his predecessor’s import-substitution policies and replacing them with reduced government subsidies, stabilized exchange rates, efforts to attract foreign investments, and development of large scale infrastructure projects. These policies were definitely important, but our argument is that these policies would have yielded much less in the absence of linguistic efficiency.

5. STATISTICAL ANALYSIS: RESEARCH DESIGN
To further understand the link between linguistic standardization and growth, here we employ a statistical test. Our sample is all countries 1978-2009. The choice of years was driven primarily by data availability for linguistic standardization. The unit of analysis is country-year. In all, we have more than 3800 observations across 155 countries. Note that some observations are dropped in various analyses due to missing data.

Dependent Variable: Economic Growth
Our primary dependent variable of interest is economic growth. Here, we use data as reported in the Penn World Table (Heston, Summers, and Aten 2009). Growth is measured as change in real GDP per capita in constant prices. The average growth rate in the entire sample is 1.61.

Key Explanatory Variable: Linguistic Standardization
Our key explanatory variable is linguistic standardization. Specifically, we are interested in the population that is proficient in the official language of the state. This population can include those who consider the language a mother tongue (L1) and those who have learned it as a second/third language (L2). Mathematically, the population of L1+L2 speakers is always at least—if not always greater—than simply the L1 population. To measure this variable, we use data from Ethnologue. Ethnologue, a database developed in 1934, is arguably today’s most comprehensive linguistic encyclopedic reference. Since the first edition, there have been sixteen subsequent editions. The most recent edition was published in 2013. The benefit of using Ethnologue is twofold. First, it identifies the number of speakers (primary and when available secondary as well) for each language. Another advantage of Ethnologue over other encyclopedic sources (e.g., the CIA World Factbook) is that it chronicles this number at multiple points in time (e.g., 1978, 1984, 1988, 1992, 1996, 2000, 2005, and 2009). The size of L1+L2 speakers of the official state language ranges from 0 to 1, where 1 indicates the entire population is proficient in the said language.

Key Alternative Variable: Ethno-Linguistic Heterogeneity
The primary alternative hypothesis we seek to challenge is about the effects of linguistic heterogeneity. To this end, we employ two different measures. The first is Alesina et al.’s
(2003) ethnic fractionalization index. The index measures the likelihood that two randomly drawn individuals are from different ethno-linguistic groups. The larger the index (maximum value=1), the more heterogeneous the country. A minimum value of 0 indicates complete ethnic homogeneity, and vice versa. Theoretical expectations would suggest more heterogeneous countries have lower levels of growth (Easterly and Levine 1997).

The alternative measure is the size of the largest ethno-linguistic group. As the size of the group increases, the extent of preference congruity, the density of networks, and the incentives to cooperate all increase as well. All else being equal, a country whose largest ethnic group is big has higher levels of growth than a country whose largest ethnic group is small. Data for this measure come from Cederman, Wimmer, and Min (2010). Note that there are two distinct differences between this Cederman, Wimmer, and Min measure and the key explanatory variable Linguistic Standardization. First, while the language of the largest ethnic group may also be the official state language, this is not always the case. The variable Linguistic Standardization focuses on the official language of the state; in contrast, the Cederman, Wimmer, and Min measure focuses on the population that speaks the language of the largest ethnic group. Second, even if the two languages are the same, the measure for Linguistic Standardization includes second-language speakers.

Control Variables
There are a number of other factors that could possibly affect a country’s growth rate. To remedy this, we include a set of control variables based on theoretical priors. First, to control for temporal autocorrelation, we include a lagged dependent variable. Naturally, we expect the coefficient for the lagged dependent variable to be positive and highly significant. Other controls are as follows:

Democracy: Regime type can be important because democracies are characterized by low barriers to entry and exit and low costs for participation (Baum and Lake 2003; Mulligan, Gil, and Sala-i-Martin 2004). This link between government and citizenry constrains the former to adopt policies that reflect the preferences of the latter. On the one hand, this link can be conducive to growth. Democracies invest more in human capital (Brown and Hunter 1999, 2004; Stasavage 2005). They are also better at protecting properties (Clague et al. 1996; Frye 2004; North and Weingast 1989; Riker and Sened 1991). On the other hand, however, the link can also mean democratic governments are more likely to be captured by particularistic pressures (Jensen 2008). This can be especially problematic if these demands are short-termed and not conducive to long term growth. To measure regime type, we employ the 21-point Polity index, where a -10 (minimum) indicates complete authoritarian and a 10 (maximum) indicates complete democracy.

British Colonialism: Another important variable is colonial legacy, specifically a British legacy (see La Porta et al. 1999). Although the British and continental Europeans both pursued assimilation in their respective colonies, the two sets of powers adopted different courses of action. Generally, the former pursued indirect rule, incorporated the locals into the state administration, and favored multilingualism in the school curriculum. This meant upon independence, British colonies were poised to inherit a set of quality of institutions that was lacking in their non-British counterparts. To control for British colonial legacy, we use data from Hadenius and Teorell (2005).

Population: We include a measure of population size to capture the size of the potential labor force. When there is a large labor force, a country can export at higher volumes.
Similarly, the availability of labor can make a country attractive to foreign investors. We use data provided by the Penn World Table (Heston, Summers, and Aten 2009). Given the non-normal distribution of this variable, we take the log.

**GDP/Capita:** We also include GDP per capita to reflect directly the wealth of the average citizen and to measure indirectly the market size. Wealthier countries attract more foreign investors, have stronger knowledge-based economies, and invest more in research-and-development. Figures come from Penn World Table (Heston, Summers, and Aten 2009) and have been rescaled to the 10,000 unit.

### 6. STATISTICAL ANALYSIS: EMPIRICAL EVIDENCE

We estimate our models using random effects GLS regression with standard errors clustered by country. The results are presented in Table 2. We begin with a baseline model that looks at the effect of ethno-linguistic fractionalization without controlling for the size of the majority or the number of L1+L2 official language speakers. The results in Model 1 are consistent with extant research. The large negative coefficient suggests heterogeneous countries have lower economic growth rates than their homogeneous counterparts. In fact, a one-standard deviation shift from below the mean (0.2044) to the mean (0.4645) in ethno-linguistic can decrease growth rates from 2.0268% to 1.4876%.

<table>
<thead>
<tr>
<th>Table 2 about here</th>
</tr>
</thead>
</table>

Among the control variables, three are significant. First, democracy, despite the theoretical impasse discussed above, has a positive effect. When governments are electorally constrained, they have an incentive to adopt growth-enhancing policies. The growth rate in the median authoritarian regime (Polity=6) is only two-thirds of that of the median democracy (Polity=9). Second, colonial legacy matters as well. This is consistent with prior expectations that the British established better institutions in their colonies. Upon independence, these states inherited administrative apparatuses, economic infrastructures, and education curriculums that were all much better equipped for promoting growth. Growth rates in British colonies are 33% higher than those in non-British colonies. Third, as would be expected, GDP growth rates are highly contingent on those from the previous year.

Model 2 replicates the previous baseline model but replaces the ethno-linguistic fractionalization measure with size of the largest ethnic group. The results remain substantively unchanged. Growth rates are higher in homogeneous countries, democracies, and British colonies. They are also higher when growth rates in the preceding year are also high. Of particular interest here is the effect of group size. When the size of the largest ethnic group increases from 62.21% (mean) to 93.17% (one-standard deviation above the mean), the growth rate corresponding increases from 1.49% to 1.82%.

The results in the first two models are consistent with the literature: Heterogeneity retards growth. The mechanism has to do with the fact that individuals from different ethno-linguistic groups are less likely to share the same preferences and are less likely to interact one another (whether by choice or by structure). In Model 3, we replace the measure for ethnic heterogeneity with one that focuses on the size of the population that speaks the official government language. The logic is that when a population is largely conversant in the same language—and that language is also that of the government—this facilitates efficiency in economic exchanges. The results lend support to this claim. The coefficient for \( L1+L2 Size \) is significant and in the positive direction (\( \beta=0.9157; SE=0.3990 \)). Substantively, as the number of people able to converse in the official state language increases, economic growth levels increase as well.
Mathematically, the effects on growth are very comparable to those of the largest ethnic group size. As a point of reference, the same shift from 62.21% to 93.17% (mean and one-standard deviation of largest ethnic group size) increases growth rate from 1.52% to 1.80%.

The results in Model 3 remain unchanged even when we take into account ethno-linguistic heterogeneity (Model 4). In fact, the magnitude for $L1+L2\ Size$ increases to $\beta=1.614$ (SE=0.6532). The same shift from 62.21% to 93.17% increases growth rate from 1.60% to 2.10%. Interestingly, the coefficient from ethno-linguistic heterogeneity loses its significance. Moreover, it is in the wrong direction ($\beta=0.8883$; SE=0.7846). Taken together, these results in Table 2 tell an important story. Despite theoretical claims that ethno-linguistic heterogeneity impedes economic growth because of inter-group tensions, we find no evidence of this effect once we take into consideration second language speakers. When people—regardless of their ethnic identity—can converse in the official state language, this reduces transaction costs in economic exchanges. This efficiency in turn facilitates economic growth.

**Discussion**

The results in Table 2 indicate size matters. Specifically, as the number of speakers of the official state language increases, growth levels increase as well. This is true regardless of whether the speaker knows the official language as a mother tongue (L1) or has learned it as a second language (L2). This suggests the link between language and growth is not always about ethnic group identity per se. Instead, language, as a vehicle of communication, can facilitate interaction in a variety of economic exchanges. When two individuals, even if coming from different ethnic backgrounds, can communicate in the absence of a translator, this reduces transaction costs.

Individuals whose mother tongue is not the official state language can learn the latter language when they are exposed to it in the classrooms. Here, the education sector is important because it is an “ideological status apparatus” (Althusser 1971)—in the government’s approved language. Regardless of the cultural narrative being taught in the classrooms (Kaplan and Baldauf 1997), students are exposed to and educated in the official state language. As more children attend school and attend for longer periods of time, the population familiar with the official state language increases. *A priori*, countries that have a large schooled population in the official state language have higher growth rates.

To test this, we look at the female secondary education enrollment rate, as a percentage of the population in the theoretical age group. Data come from the United Nations Institute for Statistics. We focus on female enrollment because any education system characterized by the general absence of girls would indicate at least half of the able-bodied workforce is not being exposed to the formal learning of the official language. Additionally, we are interested in the secondary level of education because it better captures the population of students who are exposed long term to the official language.\(^2\) Taken together, a country with a high female secondary enrollment rate would indicate an education system that focuses on all children and this attention constitutes a substantive number of years of their pre-adult lives. In contrast, a country with a low female secondary enrollment rate would indicate either (1) a gender-equal education scenario but where the boys and girls are not attending school for a substantive period of time or (2) an education curriculum that discourages girls from learning. Regardless of which of the two scenarios is at play, we argue both indicate a situation where a sizable portion of the

\(^2\) Admittedly, while enrollment in secondary education is not sufficient for literacy, the former is necessary for the latter (Birdsall, Ross, and Sabot 1997).
population (>50%) is not exposed to the official language and hence lack proficiency in the language of concern. Given the potential lagged effect of education, we lag female secondary education enrollment rates.

The results in Table 3 show the importance of education. In Model 5, we see female secondary enrollment rates have a positive and significant effect on economic growth (β=0.0261; SE=0.0129). Specifically, as the former increases from one standard deviation below the mean (33.70) to the mean (69.65), the latter correspondingly increases from 3.43% to 4.27%. These numbers are consistent with our theoretical priors. Again, it is interesting to observe that the coefficient for ethno-linguistic fractionalization, while in the right direction (negative), is not significant. This non-finding is additional evidence that the purported economic effects of ethnic tensions can be mitigated when there is a unifying language cutting across the diversity.

[Table 3 about here]

Model 6 employs an alternate measure of ethnic heterogeneity. Here, we see the size of the largest ethno-linguistic group has a significant effect (β=0.0197; SE=0.0119). Consistent with the results from Table 3, as the size of the group increases (thereby suggesting increasing homogeneity), growth levels increase as well. Additionally, the coefficient for female secondary enrollment rates is still significant. These results highlight the possibility of an interaction effect. Specifically, while homogeneity (as measured by the presence of one large ethnic group) may facilitate economic growth, the purported negative effects of heterogeneity can be mitigated by teaching the official state language to a large number of students. In Model 7 we explore this interaction.

The results in Model 7 indicate that while the two constituent terms are significant (and in the correct direction), the coefficient for the interaction term is not (β=-0.0224; SE=0.0233). However, it is important to recall that the language of the largest ethnic group may not always match the official state language. In such scenarios, even if female secondary enrollment rates are high, these students are learning not the majority language but that of the state. Hence, there is no theoretical prior to believe the interaction between heterogeneity and education would have a significant effect. But if we restricted the sample to only focus on the cases where the majority language is also the official state language, then we should see schooling to moderate the effects of heterogeneity. The results in Model 8—illustrated in Figure 2—lend support to this discussion.

[Figure 2 about here]

Three comments merit discussion. First, secondary enrollment rates have no significant effect in countries whose largest ethnic group is sizable. Second, when female secondary enrollment rates are low, countries whose largest ethnic group is big are economically stronger than their counterparts whose largest ethnic group is small. But third, and most importantly for this paper, when enrollment rates are high, the effects of group size have no bearing. In sum, this figure highlights that while homogeneity has undisputable advantages for growth, the negative effects of heterogeneity can be ameliorated with extensive exposure to the official language in schools. In fact, extensive schooling can equalize the effects of ethnic heterogeneity and homogeneity.

7. CONCLUSION
This research has important implications for both governments and scholars. For future research, it is clearly important to consider the mechanisms that links ethnic diversity and economic outcomes. This paper addresses two of the main problems with research on ethnic diversity—
conceptualization and measurement. Conceptually, language is not only an ethnic marker but also a tool for communication. Language is also additive—individuals can speak their mother tongue but also additional languages—and these multiple languages are not given at birth but can be acquired through time.

Our research also has implications for the way researchers measure diversity. Since individuals can learn languages, in particular through the education system, language should be measured as being dynamic over time and not simply static at any given point. For our research, the standardization of the language is important. We find that economic growth is affected not by diversity per se but by the proportion of the population that speaks the language of the government.

This findings paper contributes to our understanding of the effect of ethnic diversity on economic growth and to the importance of government policy for economic growth by answering two broad questions. First, contrary to conventional wisdom, heterogeneous countries are not stuck in some sort of “growth tragedy” (Easterly and Levine 1997). While they may not have the easy benefits as their homogeneous counterparts, these countries can generate similar effects through linguistic standardization. As the number of people who speak the official language of the state—whether as a mother tongue or as a second language learned in schools—increases, this has positive implications for economic growth. This is regardless of the ethnic identity of the speakers.

Second, by focusing on the effects of second languages learned, this paper finds that education in a common language can help overcome the effects of ethnic diversity and enhance economic growth. We build on the extensive literature on the relationship between ethnic diversity and economic growth by incorporating multilingual individuals and education policy. We suggest that education and the standardization of a linguistic regime allows countries to overcome the adverse effects of diversity and grow at a rate comparable to homogeneous countries. Education can increase the number of people who speak the official language and who can thus communicate with each other and with the government.
REFERENCES


TABLE 1: The Tale of Two Countries

<table>
<thead>
<tr>
<th></th>
<th>Indonesia</th>
<th>The Philippines</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colonial Ruler</td>
<td>Non-British: Dutch</td>
<td>Non-British: Spanish</td>
</tr>
<tr>
<td>Geographical Spread</td>
<td>Archipelago</td>
<td>Archipelago</td>
</tr>
<tr>
<td>Ethnic Heterogeneity</td>
<td>High (0.815)</td>
<td>High (0.845)</td>
</tr>
<tr>
<td>Economic Ideology</td>
<td>Capitalist (Non-Communist)</td>
<td>Capitalist (Non-Communist)</td>
</tr>
<tr>
<td>Political Regime</td>
<td>Authoritarian: Suharto 1968-98</td>
<td>Authoritarian: Marcos 1965-86</td>
</tr>
<tr>
<td>Linguistic Repertoire</td>
<td>Malay: Standardized</td>
<td>English: Not Standardized</td>
</tr>
<tr>
<td>Economic Growth</td>
<td>“Miracle”</td>
<td>“Sick Man”</td>
</tr>
<tr>
<td>TABLE 2: Economic Effects of Linguistic Standardization</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>DV: GDP Growth</strong></td>
<td><strong>Baseline</strong></td>
<td><strong>Full</strong></td>
</tr>
<tr>
<td></td>
<td>Model 1</td>
<td>Model 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ELF</td>
</tr>
<tr>
<td><strong>Linguistic Fractionalization</strong></td>
<td>-2.0768$^\dagger$ (0.5578)</td>
<td></td>
</tr>
<tr>
<td><strong>Majority Group Size</strong></td>
<td></td>
<td>1.0564$^\dagger$ (0.4805)</td>
</tr>
<tr>
<td><strong>L1+L2 Official Size</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Democracy</strong></td>
<td>0.0447$^\dagger$ (0.0204)</td>
<td>0.0455$^\dagger$ (0.0215)</td>
</tr>
<tr>
<td><strong>British Colonialism</strong></td>
<td>0.5009$^*$ (0.2815)</td>
<td>0.4763$^*$ (0.2861)</td>
</tr>
<tr>
<td><strong>(logged) Population</strong></td>
<td>0.1475 (0.1025)</td>
<td>0.1714 (0.1127)</td>
</tr>
<tr>
<td><strong>GDP/Capita (10,000)</strong></td>
<td>0.0262 (0.1207)</td>
<td>0.1349 (0.1233)</td>
</tr>
<tr>
<td><strong>GDP Growth (lagged)</strong></td>
<td>0.1791$^\dagger$ (0.0390)</td>
<td>0.1835$^\dagger$ (0.0390)</td>
</tr>
<tr>
<td><strong>Constant</strong></td>
<td>0.7452 (0.9738)</td>
<td>-1.2011 (1.0678)</td>
</tr>
<tr>
<td><strong>N (Groups)</strong></td>
<td>3863 (155)</td>
<td>3871 (152)</td>
</tr>
<tr>
<td><strong>Wald $\chi^2$</strong></td>
<td>93.12$^\dagger$</td>
<td>73.44$^\dagger$</td>
</tr>
</tbody>
</table>

**Note:**
Standard errors clustered by country. * $p \leq 0.100$, † $p \leq 0.050$, and ‡ $p \leq 0.010$. 
TABLE 3: Economic Effects of Language Exposure in Schools

<table>
<thead>
<tr>
<th>DV: GDP Growth</th>
<th>Baseline</th>
<th>Interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model 5</td>
<td>Model 6</td>
</tr>
<tr>
<td></td>
<td>ELF</td>
<td>Majority</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linguistic Fractionalization</td>
<td>-1.8036 (1.6277)</td>
<td>1.9744* (1.1881)</td>
</tr>
<tr>
<td>Majority Group Size</td>
<td></td>
<td>1.9744* (1.1881)</td>
</tr>
<tr>
<td>‡ 2ndary Enrollment (lagged)</td>
<td>0.0261† (0.0129)</td>
<td>0.0229* (0.0130)</td>
</tr>
<tr>
<td>Group Size * Enrollment</td>
<td></td>
<td>-0.0224 (0.0233)</td>
</tr>
<tr>
<td>Democracy</td>
<td>-0.117 (0.0804)</td>
<td>-0.123 (0.0817)</td>
</tr>
<tr>
<td>British Colonialism</td>
<td>-0.3703 (0.8104)</td>
<td>-0.4063 (0.8094)</td>
</tr>
<tr>
<td>(logged) Population</td>
<td>0.052 (0.2048)</td>
<td>0.0528 (0.2136)</td>
</tr>
<tr>
<td>GDP/Capita (10,000)</td>
<td>-0.5565* (0.3084)</td>
<td>-0.4643 (0.2836)</td>
</tr>
<tr>
<td>GDP Growth (lagged)</td>
<td>0.1554* (0.0912)</td>
<td>0.1578* (0.0915)</td>
</tr>
<tr>
<td>Constant</td>
<td>3.3085 (2.3230)</td>
<td>1.3307 (2.2557)</td>
</tr>
<tr>
<td>N (Groups)</td>
<td>355 (130)</td>
<td>358 (131)</td>
</tr>
<tr>
<td>Wald χ²</td>
<td>7.79</td>
<td>8.30</td>
</tr>
</tbody>
</table>

Note:
Standard errors clustered by country. * p ≤ 0.100, † p ≤ 0.050, and ‡ p ≤ 0.010.
FIGURE 1: Economic Growth in Indonesia and the Philippines 1978-2007

GDP Growth %

Indonesia  Philippines

78  81  84  87  90  93  96  99  02  05

19
FIGURE 2: Conditional Effects of Ethnic Heterogeneity and Enrollment Rates

Majority Size: 0.33  Majority Size: 0.67

Female Secondary Education Enrollment Rate

GDP Growth %