

Weingarten, Severin; Uebelmesser, Silke

**Conference Paper**

## Language Learning and Migration: A New Dataset on Language Course and Exam Participation

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2015: Ökonomische Entwicklung - Theorie und Politik - Session: Migration, No. G05-V1

**Provided in Cooperation with:**

Verein für Socialpolitik / German Economic Association

*Suggested Citation:* Weingarten, Severin; Uebelmesser, Silke (2015) : Language Learning and Migration: A New Dataset on Language Course and Exam Participation, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2015: Ökonomische Entwicklung - Theorie und Politik - Session: Migration, No. G05-V1, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft

This Version is available at:

<https://hdl.handle.net/10419/112958>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# Language Learning and Migration: A New Dataset on Language Course and Exam Participation

Early draft (Do not cite!)

March 1, 2015

## **Abstract**

Optimal integration policies should take the language skills of arriving migrants into account. However, little is known about the determinants of these skills, because most survey-based studies cannot distinguish between language learning before and after migration. We present a new dataset which reports the extent of language course and exam participation at the German Goethe institutes in 91 countries for the period 1966–2013. The dataset allows for a detailed examination of the determinants and consequences of adult language learning on the institute and country level. We estimate a fixed-effects model for the relationship between language learning and a set of macro-economic variables. Immigration flows are positively correlated with exam participation. Additionally, in EU countries, migrant stocks are positively correlated with course and exam participation. They may act as a proxy for otherwise unobserved short-term migration and cultural ties, both of which could influence language learning decisions. In non-EU countries, the link between migration and language learning is strengthened by two factors: By positive economic conditions, indicating that migrants who leave amidst economic turmoil may arrive with worse language skills. And by linguistic distance, indicating that migrants react to distance by increasing learning effort instead of refraining from learning the language.

*JEL classification:* F22, J24, J61.

*Keywords:* language skills, language learning, international migration, migrant networks, migration policies.

# 1 Introduction

Language skills play an important role in shaping international migration flows and in determining if and how migrants integrate into their host societies. We present a new dataset which reports the extent of language course participation at the German Goethe institutes in 91 countries for the period 1966–2013. The dataset allows for a detailed examination of the determinants and consequences of adult language learning on the macro-level. In this paper we describe the dataset. Additionally, in a first exploratory exercise, we relate language course participation to several key variables, most importantly migration, economic conditions, trade, and linguistic distance.

The literature on migration choice has long recognized the importance of language skills by controlling for common languages of origin and destination countries (e.g. Mayda 2010; Grogger and Hanson 2011; Belot and Hatton 2012; Ortega and Peri 2013). More recently, Adsera and Pytliková (2012) and Belot and Ederveen (2012) show that linguistic distance, which can be interpreted as the difficulty associated with learning another language, has an important effect on international migration flows.<sup>1</sup> Aparicio Fenoll and Kuehn (2014) find a strong positive effect of school-age language learning on migration between EU countries. Their results show that the study of language learning processes can add value to a literature which has previously focused on linguistic properties. While linguistic properties are beyond the reach of policy makers, language learning is not. It can be part of school curricula, but it can also be encouraged or even made a requirement by the governments of destination countries.

In addition to their importance as a determinant of migration, language skills are also crucial for the integration of migrants into their host societies. They have a positive effect on earnings (Chiswick and Miller 1995; Dustmann and Soest 2001; Dustmann and Fabbri 2003; Bleakley and Chin 2004), employment (Dustmann and Fabbri 2003), and social assimilation (Bleakley and Chin 2010), and they can affect occupational choice (Chiswick and Miller 2007).

Given the importance of language skills, a large number of studies has explored their determinants. Chiswick and Miller (2014, section 4) offer an extensive review of the literature where they group determinants into three categories, which they dub “the three E’s”: exposure, efficiency and economic incentives. Exposure depends on time since migration, residence in ethnic enclaves and the language spoken by spouse and children. Efficiency variables include age at migration, level of education, linguistic distance, and the

---

<sup>1</sup>While this paper focuses on international migration, the effect of linguistic barriers is not limited to the factor labor. Lohmann (2011) and Ispording and Otten (2013) show that linguistic distance also has a negative effect on trade.

motive of migration. Two economic incentives for language learning that are addressed in the literature are expected duration of stay and expected gains in earnings from language proficiency.

Since most of these variables vary on the level of the individual migrant, studies on the determinants of migration use censuses or surveys to obtain micro-level data. Typically, these datasets report language skills that respondents possess at the time of the collection of the data. However, the timing of language learning is relevant in its own right. Foreign language acquisition at early ages occurs primarily at school and is determined by the schooling system. For adults on the other hand, the decision to learn a language is more likely to be made in light of a decision to migrate. Migrants who possess language skills at some point in time after their arrival in the host country may have been selected on the basis of pre-existing skills or they may have been motivated to learn the language by their decision to migrate. To the best of our knowledge non of the micro-level studies on the determinants of language skills can distinguish between these two effects. However, from the point of view of the policy maker, an understanding of the motivation effect is highly relevant, because it allows the targeting of language courses at groups of immigrants who are more likely to lack necessary language skills. While the dataset presented in this paper does not contain micro-level information about individual migrants, it allows for an explicit focus on the motivation effect because it reports on language learning and not on the presence of language skills.

The rest of the paper is structured as follows. Section 2 introduces the new dataset. Section 3 outlines some preliminary hypotheses which are examined using the empirical setup presented in section 4. We discuss our first results in section 5 and provide an overview of our next steps in section 6.

## 2 The Dataset

The Goethe-Institut (GI) is a German association which promotes the study of the German language and the German culture abroad. Most importantly for our purposes, it maintains more than 150 institutes in 90 countries, at many of which locals can study the German language and obtain language certificates which are widely recognized. The GI is mainly funded by the German government and through course fees (Goethe-Institut e.V. 2013).

The dataset presented in this paper contains yearly observations for each institute in the period 1966–2013. It covers four variables: the average number of course participants per course term,<sup>2</sup> the total number of course hours taught, the number of language exam

---

<sup>2</sup>The length of course terms differs across institutes. The average number of participants per course term can be thought of as the number of students who are currently enrolled in a language course at any

participants, and the number of language teachers employed. Some of the variables are not available for all years due to changes in the reported statistics, but many gaps can potentially be filled on the basis of internal records kept by the institute. Table 1 provides an overview of some descriptive statistics for each variable.

Table 1: Descriptive Statistics GI Dataset<sup>1</sup>

Variable	Interval	Min.	Median	Mean	Max.
Course hours	1966–2013	10	4390	6115	38437
Exams	1990–2013	1	110	540	26159
Students	1966–1999 <sup>2</sup>	2	427	602	6314
Teachers	1966–1999	1	4	6	36
Institutes	1966–2013	99	111	112	125

<sup>1</sup> At the time of submission of this draft of the paper we have not fully digitized all parts of the dataset. Specifically, the years 1966–1971 are still missing and no plausibility checks have been done for the variable ‘Teachers’.

<sup>2</sup> While the students variable is no longer reported in the yearbooks after 1999, it is available in internal documents provided by the GI and we plan to extend it to 2013.

Most of the data was digitized from the yearbooks of the GI which have been published continuously since 1966. The yearbooks contain detailed reports on the activities of the GI, as well as key statistics for the entire association and for each institute. The data was typed into CSV files.<sup>3</sup> Country and city names were harmonized and matched to codes used by the UN population division to allow merging with other country-level and city-level datasets.

Figure 1 plots the development of the aggregates of all four variables and the number of institutes that offer language courses over time. The number of institutes is fairly stable, with a small increase at the end of the 1970s and a small decrease in the 1990s. Course hours increase steadily throughout the entire period of observation, almost doubling from 500,000 in 1972 to 1 million in 2013. Student numbers are currently only available until 1999. They are fairly stable, but there are three periods of relatively small temporary increases at the beginning of the 1970, around 1980 and around 1990. Two potentially related developments come to mind: First, all three increases coincide roughly with stretches of relatively strong German GDP growth. Second, the largest of the three increases, around 1990, might reflect increased interest in the German language at the end of the cold war.

The number of teachers decreases slightly during the 1990s, but this variable is not yet available for the entire period of observation (see above). Language exam numbers are only available from 1990 onwards and increase steadily until 2006. They then experience a point in time during the year.

<sup>3</sup>We are very much indebted to Maik Wehlte for doing a lot of digitization work and to Martin Ahmad, Toni Grimm, and Lars Other, who helped with the digitization of additional data and plausibility checks.

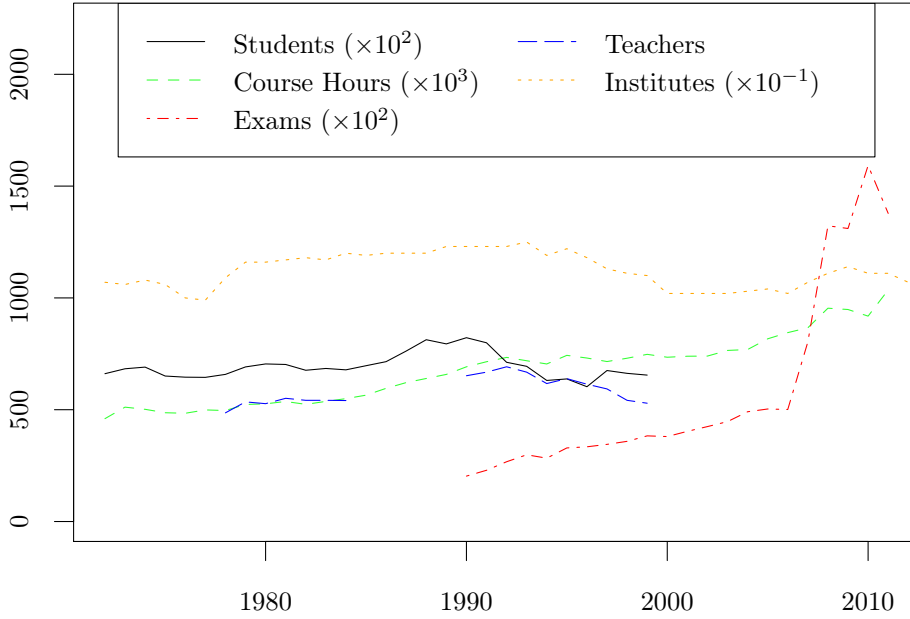


Figure 1: Aggregate time series of variables in GI dataset

strong surge to over 150,000 in 2010, more than three times the level of 2006. This surge could be related to a change in German immigration law. Since 2007, migrants who come to Germany under family reunification provisions (from non-EU countries) have to provide evidence that they possess basic German language skills<sup>4</sup>.

For the rest of this paper, we will focus our attention on exam participant and student numbers because they are the best measure of migrants language learning efforts in the dataset. The alternative measure, course hours, does not take into account, how many students attend a course. Figure 2 plots exam participant and student numbers for four selected institutes, two in EU countries and two in non-EU countries. Two observations can be made regarding the nature of the variance in the data.

First, the average numbers of exam participants and students differ considerably between institutes and these differences cannot be explained by differences in population alone. While the size of the catchment area of an institute is not obvious a priori, neither city population nor country population can explain the difference in exam participant or student numbers for Porto and Amsterdam. For example, Porto has about three times as many language students, but the metropolitan areas of Amsterdam and Porto are roughly the same in terms of population size and Portugal is considerably smaller than the Netherlands. In an analogous comparison, differences in population cannot explain the fivefold difference between average exam participant and student numbers in Ankara

<sup>4</sup>Equivalent to the A1 level in the Common European Framework of Reference for Languages (CEFR)

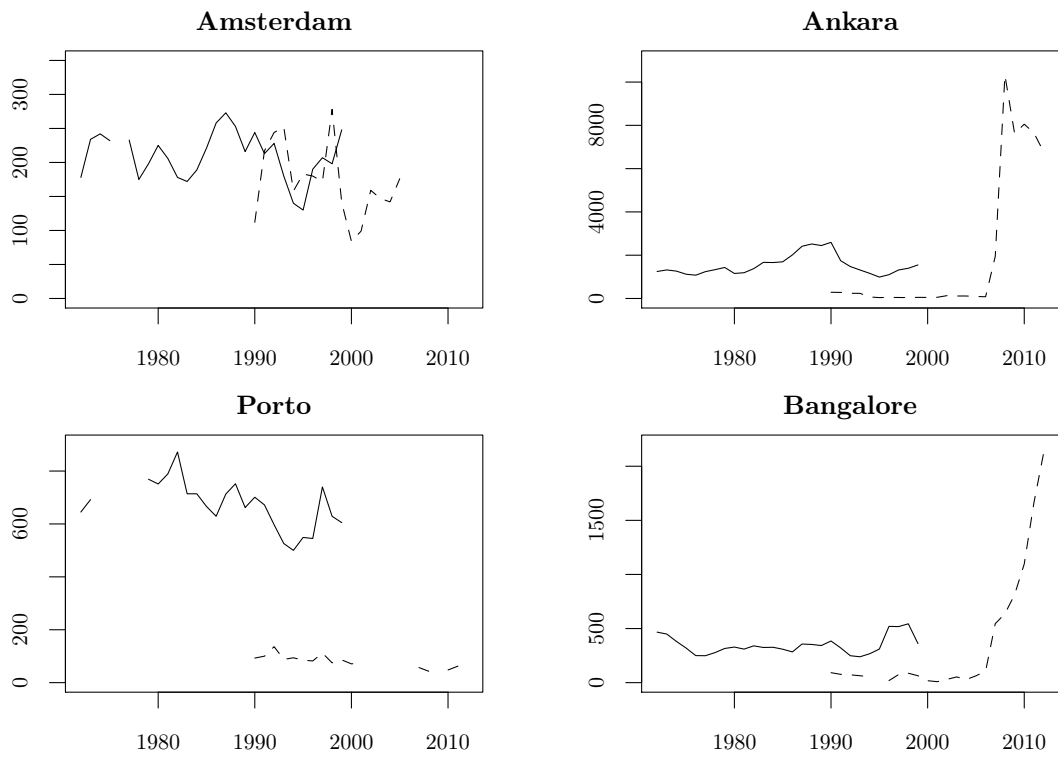


Figure 2: Development of exam participant (dashed) and student numbers (solid) for selected institutes

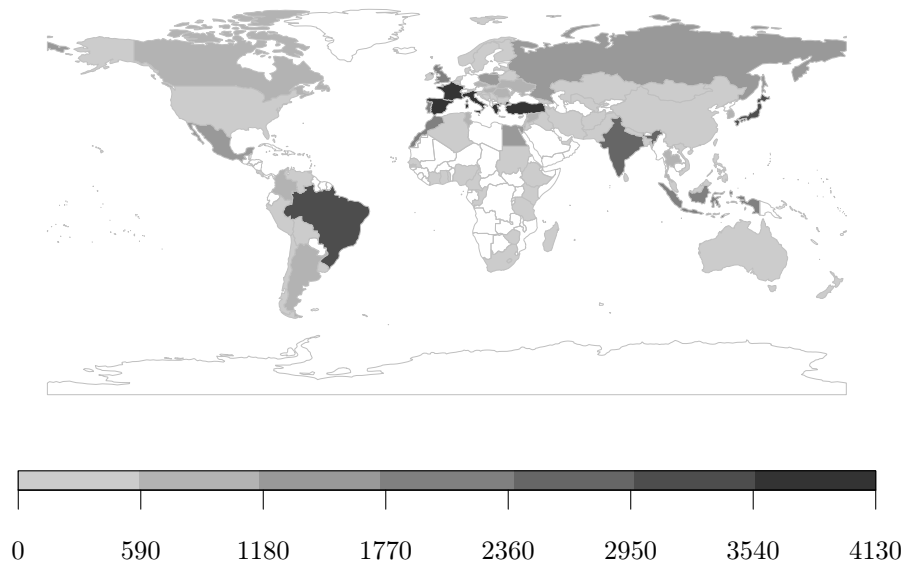


Figure 3: Student numbers at Goethe institutes in 1999

and Bangalore. A similar picture emerges from figure 3. While some of the countries with a lot of students are among the most populous in the world (e.g. India, Brazil), others are a lot smaller (e.g. South Korea, Greece).

Second, for each institute, both measures vary considerable over time. This variation is not strongly correlated between institutes. While Ankara sees a considerable increase in student numbers in the late 1980s, this increase is much weaker in Amsterdam and Bangalore and barely happens in Porto. Student numbers in Amsterdam and Bangalore increase considerably in the late 1990s, but no similar increase happens in Porto and Ankara. This suggests that student numbers are not only driven by a general trend or changes in Germany, but by country-specific factors. We will attempt to explore these factors in the following.

### 3 Hypotheses

In this section, we present several hypotheses regarding potential determinants of language learning. The first set of hypotheses relates language learning to migration and the second one relates language learning to trade.

In our estimation, we use language course and exam participation at Goethe institutes as a proxy for total language learning in the respective countries. Naturally, there are a large number of other language learning opportunities, including universities, private language schools, and internet platforms. The multitude of other options gives rise to a number of concerns regarding the self-selection of language learners into courses offered by the Goethe institutes. Three characteristics on which self-selection may be based are willingness or ability to pay, location, and age:

Selection on willingness to pay could occur if the prices of courses at the Goethe institutes differed significantly from the costs of other equally suitable learning options. On the one hand, one might suspect the Goethe institutes to be somewhat of a premium provider of language courses, because they are a semi-official German organization with a long tradition and a good reputation. Such a status would allow them to charge higher prices. On the other hand, one might suspect Goethe courses to be particularly cheap, because the majority of the Goethe institutes' funds comes from the German government.<sup>5</sup> Historical price data on language courses are not available to the best of our knowledge. However, table 2 contains current price data on comparable language courses offered by the Goethe institutes and by other institutes in six cities in different countries.<sup>6</sup> While the data are

---

<sup>5</sup>Several employees of the Goethe institute have stated in conversations with us that their language courses are priced to be self-financing and that government funding is used for other cultural activities.

<sup>6</sup>Data comes from the websites of the course providers. The websites of non-Goethe-institute providers



City	Provider	Course Type	Price / Hour	Currency
Mexico City	GI	Extensive	146.67	MXN
Mexico City	Tecnológico de Monterrey	Extensive	90.91	MXN
Buenos Aires	GI	Extensive A1.1	80	ARS
Buenos Aires	Sprachzentrum Buenos Aires	Extensive A1.1	65	ARS
Rio de Janeiro	GI	Extensive A1	56.21	BRL
Rio de Janeiro	Baukurs	Extensive A1	49.17	BRL
Lissabon	GI	Extensive	5.67	EUR
Lissabon	ilnova	Extensive	6.17	EUR
Ankara	GI	Extensive A1	10.16	TRY
Ankara	Hitit Education Institutions	Extensive A1	11.88	TRY
Tokyo	GI	Intensive	1541.67	JPY
Tokyo	German Office	Intensive	1971.43	JPY

Table 2: Prices of language courses at Goethe institutes and other providers in 2015

far from complete or representative, they do not indicate that the Goethe institutes are usually the most expensive provider in the market.

Goethe institutes are usually located in capitals and other major cities. The lack of institutes in rural areas is likely to lead to an underrepresentation of language learners from these areas among participants at the Goethe institutes. However, the bias need not be as large as one would initially expect: Goethe institutes offer both extensive and intensive language courses. Extensive courses are based on weekly lessons and last for several months, but intensive courses are taught en-block. Participants of intensive courses do not necessarily have to live in the vicinity of the respective institute. They may also stay there for the duration of the course only.

The language courses taught by the Goethe institutes are a traditional “offline” form of language learning. At the other end of the spectrum are pure online courses like those offered by “myngle” or “babble”. The latter kind of courses may be more attractive to a younger generation of language students, which is more familiar with using the internet in general. While this difference may lead to an overrepresentation of older students among the participants in language courses at the Goethe institute, the advent of online language learning platforms in the late 2000s falls in the very last years of our period of observation from 1980 to 2009. Consequently, we assume that the age bias has only a small, if any, effect on our results.

Independently of the question of the representativeness of the GI data, language learning

---

were found by searching Google for “language learning” and the name of the respective city in the native language of the respective country.

may generally be less common among very high-skilled migrants who speak English on the job and whose social circle consists mainly of other high-skilled migrants and Germans, who are also used to communicating in English. As a consequence, we expect our results to be less representative of migrants with a very high skill level.

### **3.1 Immigration**

Given the large number of potential benefits of language proficiency and the robust results regarding the effect of proficiency on earnings, migrants should have an interest in obtaining language skills. A positive correlation between immigration and course and exam participation would be in line with the existence of an effect of migration on language learning rather than a pure selection effect based on skills acquired earlier in life.

**Hypothesis 1a** Language course and exam participation in a country is positively correlated with immigration from that country to Germany.

#### **Minority Language Concentrations**

Migrant networks are often considered to improve the ability of migrants to find work in their host country and build social ties to others who speak their native language. As a consequence, speaking the host country's language may be less important for migrants who can rely on migrant networks. Several studies find that minority language concentrations are associated with lower levels of language proficiency (Chiswick and Miller 2007; Espenshade and Fu 1997; Lazear 1999; Isphording and Otten 2013). We hypothesise that these results are not exclusively based on the (self-)selection of migrants with worse language skills, but that the negative effect of minority language concentrations extends to the language learning decisions of immigrants. We use the number of citizens of a country of origin who live in Germany as a proxy for the size of the respective minority language concentration.

**Hypothesis 1b** Language course and exam participation in a country is less strongly correlated with migration flows if a large number of migrants from that country live in Germany.

#### **Immigration and Economic Conditions**

There are at least three potentially viable channels through which economic conditions could moderate the relationship between migration and language learning: a job-search effect, a preparation-time effect and an affordability effect.

First, migrants who leave under adverse economic conditions in their home country are more likely to be economic migrants, who seek better employment opportunities abroad,

than migrants who leave when the economy is going well. If language skills increase the likelihood of finding employment, economic migrants may have a larger incentive to participate in language courses before their departure. Second, migrants who leave under better economic conditions may have more time to prepare their immigration, because they are more likely to be in employment. Migrants who leave in times of economic turmoil may be out of employment and may want to leave more quickly to find work in Germany. This would give them less time to prepare and, more specifically, to participate in language courses. Third, language courses are not free. Migrants may simply be more willing to participate in language courses if they have more money to spend.

While the job-search channel points towards a negative effect of economic conditions on the link between course participation and migration, the preparation-time and affordability channels point towards a positive effect. We use the latter as a working hypothesis. While we cannot disentangle the three channels empirically, we can tell if the aggregate effect of the three channels is positive or negative.

**Hypothesis 1c** Lower unemployment and higher GDP per capita (relative to its trend) strengthen the positive relationship between migration and course and exam participation.

### **Linguistic Distance**

The linguistic distance between two languages can be interpreted as the difficulty encountered by a native speaker of one of the languages when learning the other. As such, linguistic distance may have an effect on the language learning decision of migrants, but the direction of this effect is not clear. On the one hand, the increased effort associated with learning a more distant language may translate into more course participation. On the other hand, it may discourage language learning in the first place. Given the large benefits associated with speaking the host country's language, we would expect the former effect to dominate the latter.

**Hypothesis 1d** Greater linguistic distance strengthens the positive relationship between migration and course and exam participation.

## **3.2 Trade**

While the present paper focuses more on the importance of language learning in the migration context, language is also relevant for trade relationships. Using different measures of linguistic distance, both Lohmann (2011) and Isphording and Otten (2013) find that linguistic distance has a negative effect on bilateral trade. If language barriers can hinder trade, trade partners should have an incentive to learn each others languages. Therefore, we would expect language learning to be positively related with trade.

**Hypothesis 2a** Language course and exam participation is positively correlated with trade flows.

The link between language learning and trade, may also be affected by linguistic distance. As in the migration context, the direction of the effect is unclear. Greater linguistic distance increases the necessary effort associated with learning the trade partner’s language. The increased effort requirement may translate into more course participation, but it may also discourage learning. Again, we use the positive effect as a working hypothesis, but we suspect the relative costs and benefits of language learning to strongly depend on the specific properties of the trade relationship in question.

**Hypothesis 2b** Greater linguistic distance strengthens the positive relationship between trade and course and exam participation.

## 4 Empirical Setup

We use OLS fixed-effects regressions to estimate the effect of a set of explanatory variables on exam participation and course participation at the level of individual Goethe institutes. This gives rise to a two-level geographical structure, where most explanatory variables are available on the country-level, but where each country may contain several institutes, for each of which we observe different course participation numbers. We use institute-level rather than country-level estimations, because it allows us to exploit more of the variance in our explained variable. This approach also avoids the problem of changes in aggregate numbers when institutes open and close.

We estimate

$$P_{ijt} = \alpha + \beta x_{jt} + \delta_{Dt} D_t + \delta_{Di} D_i + u_{ijt} \quad (1)$$

where  $P_{ijt}$  is exam participation or course participation in institute  $i$ , in country  $j$  at time  $t$ .  $x_{jt}$  is a vector of country-level variables including population, migrant stocks, immigration, trade, deviation from the GDP trend, unemployment, linguistic distance, and dummies for EU and Schengen area membership.  $D_t$  and  $D_i$  are sets of year and institute dummies. The institute-level dummies capture both country-fixed and institute-fixed effects.  $\alpha$  is an intercept and  $u_{ijt}$  is an error term.

Table 3 lists all sources from which we take data for our estimations. We use gross immigration flows rather than net migration, because the latter measure contains return migration and emigration, which should not have an effect on language learning in the home country. We proxy trade flows by dividing total trade revenues with Germany by the GDP of the country in question.

Variable	Source of Data
Language course participants	Digitized Goethe institute yearbooks
Exam participants	Digitized Goethe institute yearbooks
Population	UN World Population Prospects
Unemployment	IMF World Economic Outlook (October 2013)
GDP per capita	Penn World Table 7.1
Migrant stock in Germany	German Federal Statistical Office
Immigration to Germany	German Federal Statistical Office
Trade flows	German Federal Statistical Office
EU membership	Self-compiled
Schengen area membership	Self-compiled
Linguistic distance	Melitz and Toubal (2012) dataset

Table 3: Variables and data sources

## 5 Preliminary Estimation Results

In this section, we present the results of fixed-effect regressions explaining the number of language exams taken for 1990–2009 and the number of language students for 1980–1999.<sup>7</sup> Both dependent variables come from the GI dataset described above.

Exam participation is our preferred dependent variable. Student numbers are potentially censored from above, because institutes may run into capacity constraints in periods of high demand for language courses. Exam participation is not subject to the same constraints, because exams require much less effort by the institutes. Additionally, exam participants must not be students of the Goethe institutes, making exam participation a better proxy of adult language learning in general.

We split our data into an EU and a non-EU sample, because we expect the potential benefits of language learning to differ significantly for residents of EU and non-EU countries. For example, since Germany uses an employer-based immigration system, economic migrants from non-EU countries have to find employment before moving to Germany, while EU citizens can migrate first and look for a job after arrival. If pre-migration language learning facilitates job search, it becomes less relevant as a country joins the EU. On a similar note, learning German may be more of a long-term investment for EU citizens who will almost certainly be allowed to migrate to Germany whenever they want to, while language learning will be tied to more specific migration opportunities for non-EU citizens.

We use data from all countries for which the necessary data is available for at least 75% of observations. Our final EU sample contains 28 institutes in 8 countries and our

<sup>7</sup>While the Goethe dataset starts in 1966, unemployment data is used in many of our estimations and it is only available for a large number of countries from 1980 onwards.

final non-EU sample contains 48 institutes in 27 countries.<sup>8</sup> We center all non-dummy variables, so that the main coefficient of each variable that is also used in interactions can be interpreted as an average marginal effect.

	ex-EU-1	ex-EU-2	ex-EU-3	st-EU-1	st-EU-2	st-EU-3
Population ( $\times 10^6$ )	75.77*** (12.41)	51.41*** (15.95)	21.38 (19.24)	3.75 (13.11)	-3.82 (17.34)	13.87 (21.62)
Pop. $\times$ Dev. from GDPpc	-5.42 (11.26)	-2.34 (11.22)	-60.40*** (20.55)	-19.57* (11.59)	39.45** (16.85)	60.73*** (20.18)
Pop. $\times$ Unemployment	-0.10 (0.36)	0.07 (0.37)	-0.88* (0.50)	-0.88*** (0.30)	0.33 (0.38)	0.48 (0.48)
Log Mig. Stock ( $\times 10^6$ )		375.33*** (142.99)	132.10 (177.86)		386.65*** (105.93)	484.90*** (111.24)
Immigration (t+1) ( $\times 10^3$ )		6.39*** (2.45)	5.19 (4.07)		2.41 (2.12)	4.12 (3.46)
Pop. $\times$ Trade share with Ger.		0.17 (0.17)	0.34* (0.18)		-0.73** (0.29)	-0.72** (0.32)
Imm. (t+1) $\times$ Log Mig. Stock			-2.49 (3.26)			3.40 (2.22)
Imm. (t+1) $\times$ Dev. from GDP trend			190.11*** (56.81)			-55.28* (28.50)
Imm. (t+1) $\times$ Unemployment			2.68*** (0.99)			0.00 (0.67)
Imm. (t+1) $\times$ Linguistic Dist.			0.71 (0.50)			-0.54 (0.40)
Pop. $\times$ Trade $\times$ Linguistic Dist.			0.01 (0.02)			-0.04** (0.02)
R <sup>2</sup>	0.66	0.68	0.69	0.94	0.94	0.94
Adj. R <sup>2</sup>	0.61	0.63	0.64	0.93	0.93	0.94
Num. obs.	417	417	417	472	472	472

\*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$

Table 4: Regression results for the EU sample

Table 4 presents estimation results for our EU sample. As indicated by the column labels, the dependent variable for the first three columns is *exam* participation and the dependent variable for the last three columns is *students*.

In all specifications, we control for population, deviation from GDP per capita trend, and unemployment. The latter two controls and our trade variable enter through interactions with population. The effects of these three variables on the absolute number of participants in exams or courses will necessarily depend on the size of the group that is affected by them.<sup>9</sup>

In “ex-EU-2” and “st-EU-2”, we add our migration and trade variables. We log migrant

<sup>8</sup>We omit Greece, Turkey and the US from our samples, because they are strong outliers with respect to at least one of our variables of interest and could potentially bias our OLS results: Greece, historically, has an exceptionally large number of exam participants. Turkey is the only major non-EU country with very large migrant stocks in Germany. The US send an unusually large number of immigrants to Germany. We suspect that this is due to the more than 100,000 US soldiers stationed in bases across Germany.

<sup>9</sup>To illustrate this, assume that unemployment increases by 1% in a large and in a small country. If unemployment has the same effect on language learning in both countries, the absolute number of language course participants will increase more strongly in the large country, because a larger absolute number of people is affected by the increase in unemployment.

stocks, because they include a number of large outliers to which our OLS estimation could be sensitive. We take immigration flows from  $t + 1$  rather than  $t$ , because language course participation in the home country will necessarily precede migration, even if it is a consequence of the intention to migrate.

We find a positive coefficient for migration inflows, but it is only significant for exam participation. This result provides some support for our hypothesis H1a. However, in addition to migration flows, migrant stocks seem to have a considerable influence on language learning, too. There are at least three potential channels through which migrant stocks may have an effect: First, they may act as a proxy for (repeated) short-term migration, which is not counted in official migration statistics, but still stimulates language learning. Similarly, they may act as a proxy for migration flows which occur at a different period than  $t + 1$ , and lastly, migrant stocks may lead to an increased interest in German culture in the home countries of migrants.

Adding interactions terms to test our other hypotheses (“ex-EU-3” and “st-EU-3”), the results for exam and course participation diverge: For exam participation, we find positive coefficients for both interactions between GDP and unemployment on the one hand, and migration flows on the other hand. According to our hypothesis H1c, these effects should point in opposite directions, because better economic conditions are generally associated with higher GDP, but lower unemployment. In our estimation, the positive GDP coefficient is consistent with an affordability effect, where fewer migrants learn the language if they cannot afford to do so. The positive unemployment coefficient could be evidence of the job-search effect discussed above. However, we find an almost opposite result for student numbers, where the coefficient for the interaction of migration and GDP is negative and the one for the interaction with unemployment is zero. While economic variables seem to have an effect on the relationship between language learning and migration, our hypotheses do not seem to capture this effect properly.

The same is true for our trade-related hypotheses (H2a, H2b). While trade seems to have a positive effect on exam participation, it has a negative effect on course participation. This effect becomes even more negative as linguistic distance between Germany and the trade partner increases. One potential explanation of this finding is that trade with Germany may increase at the same time as trade with other countries. Assuming that English is the lingua franca of international trade, more trade would increase the incentive to learn English instead of German.

In summary, our results for the EU sample provide some support for a general, positive relationship between migration and language learning (H1a), but they also suggest that growing migrant stocks increase language learning—either as a proxy for unobserved

migration flows or directly through increases in cultural ties. Regarding variables that may affect the link between language learning and migration, we find no evidence of an effect of ethnic enclaves (H1b), mixed evidence for an effect of economic conditions (H1c) and no evidence of an effect of linguistic distance (H1d). We also find mixed evidence for an effect of trade on language learning (H2a, H2b).

	ex-nEU-1	ex-nEU-2	ex-nEU-3	st-nEU-1	st-nEU-2	st-nEU-3
Population ( $\times 10^6$ )	3.97*** (1.36)	11.69*** (1.47)	9.83*** (1.43)	-4.55** (2.06)	-6.30* (3.49)	-9.47*** (3.43)
Pop. $\times$ Dev. from GDPpc	-3.05 (2.10)	0.59 (2.00)	-0.96 (2.13)	0.85 (2.40)	2.22 (2.58)	-11.37*** (3.41)
Pop. $\times$ Unemployment	-0.23*** (0.05)	-0.11** (0.05)	0.02 (0.06)	0.17* (0.10)	0.24** (0.11)	0.08 (0.13)
Log Mig. Stock ( $\times 10^6$ )		-203.46*** (54.90)	-206.16*** (55.21)		99.85 (91.80)	119.61 (112.87)
Immigration (t+1) ( $\times 10^3$ )		106.16*** (15.76)	41.62** (18.84)		43.60 (29.03)	17.37 (37.71)
Pop. $\times$ Trade share with Ger.		-0.43*** (0.04)	-0.22*** (0.06)		-0.09 (0.15)	-0.69*** (0.22)
Imm. (t+1) $\times$ Log Mig. Stock			-2.66 (16.29)			22.27 (20.67)
Imm. (t+1) $\times$ Dev. from GDP trend			313.70*** (66.64)			721.32*** (174.22)
Imm. (t+1) $\times$ Unemployment			0.15 (1.96)			6.80 (4.33)
Imm. (t+1) $\times$ Linguistic Dist.			9.19*** (1.62)			0.62 (3.15)
Pop. $\times$ Trade $\times$ Linguistic Dist.			-0.07*** (0.01)			0.15*** (0.03)
R <sup>2</sup>	0.58	0.69	0.74	0.79	0.79	0.80
Adj. R <sup>2</sup>	0.53	0.65	0.70	0.76	0.77	0.78
Num. obs.	650	597	597	736	736	736

\*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$

Table 5: Regression results for non-EU countries

We now turn to the estimation results for our non-EU sample in table 5.

Again, we find a positive relationship between migration and language learning, which is only significant for exam participation, but not for course participation (H1a). While migrant stocks had a positive effect on both measures of language learning in EU countries, they now have a negative effect on exam participation. While the negative effect is somewhat puzzling, the lack of a positive effect is in line with the suspicion that migrant stocks may act as a proxy of unobserved short-term migration in EU countries. This kind of migration, and with it its effect on language learning, is much less likely without laws that allow for free movement of short-term migrants across borders.

The results regarding variables that may affect the relationship between language learning and migration are more clear cut: With both dependent variables we find a significant, positive coefficient for GDP, which could be an indication that the preparation-time and affordability effects are dominating a potential job-search effect (H1c). We also find a positive coefficient for linguistic distance, supporting the hypotheses that migrants are more



likely to learn German—or invest more effort in doing so—if their native language is more distant from German (H1d). However, this effect is insignificant for language students as a dependent variable. On the one hand, this may be due to the fact that exam participation covers a larger group of language learners. Exam participants may have learned the language elsewhere. On the other hand, migration—and with it, language learning—across larger linguistic and geographic distances may have become more widespread over the last thirty years. Since the exam participation dataset covers a more recent period of observation, it would pick up the effect or more strongly.

As with the EU sample, we do not find evidence of an effect of migrant stocks on the link between language learning and migration (H1c) and the trade coefficients are either insignificant or have an unintuitive negative sign (H2a, H2b).

## 6 Conclusions and Next Steps

In this paper, we introduce a new dataset covering language course participation at Goethe institutes in 91 countries for the period 1966–2013. Four variables, students, exams, course hours and teachers, are observed yearly for almost all institutes where language courses are offered. To the best of our knowledge this is the first large-scale dataset on adult language learning. We use the former two variables for our estimations, because we consider them to be good proxies for language learning in the countries where the institutes are located. Both measures vary considerably between and within institutes and countries and this variance is not explained by differences in population alone.

	Hyp.	ex-EU	st-EU	ex-nonEU	st-nonEU
Migrant Stocks		+	+	–	0
H1a: Immigration	+	+	0	+	0
H1b: ... $\times$ Migrant Stocks	–	0	0	0	0
H1c: ... $\times$ Economic Conditions	+	mixed	–	+	+
H1d: ... $\times$ Linguistic Distance	+	0	0	+	0
H2a: Trade	+	+	–	–	–
H2b: ... $\times$ Linguistic Distance	+	0	–	–	–

Table 6: Directions of effects in hypotheses and estimation results (specifications 2 and 3)

We use fixed-effects models to estimate the relationship between language learning and several potential macro-level determinants. Our estimations cover the the periods 1980–1999, with language course participation as the dependent variable, and 1990–2009, with language exam participation as the dependent variable. Our results are summarized in table 6.

Language learning is (on average, i.e. at the mean of all interacted variables) positively correlated with immigration. This result supports our hypothesis H1a, but it only holds if language learning is proxied by exam participation. There are at least two potential explanations for our failure to find a similar result for student numbers: First, student numbers may not react sufficiently to increased demand for language courses, because capacity constraints prevent the institutes from offering additional courses. Second, exam participation may cover a wider range of learning opportunities, because those who take an exam at the Goethe institute may have learned the language in any number of ways.

Our remaining migration-related hypotheses concern the effect of other variables on the link between migration and language learning. We find mixed coefficients for economic conditions for EU countries, but clearly positive, significant coefficients for non-EU countries. We interpret these positive coefficients as evidence of an affordability effect, where migrants' propensity to participate in language courses increases with their ability to pay, or a preparation-time effect, where migrants have more time to prepare for migration under favorable economic conditions (H1c). For the more recent exams dataset we also find a positive effect of linguistic distance in non-EU countries, indicating that migrants from linguistically more distant countries of origin are more likely to learn German or invest more effort to do so (H1d). We do not find evidence of a dampening effect of migrant stocks on the link between migration and language learning (H1b).

Additionally, we investigate the link between trade and language learning, but we find puzzling, negative effects, which are often strengthened by linguistic distance.

We plan several next steps. First, we want to use time-series techniques to account for a more comprehensive set of lags of our migration-related explanatory variables. Second, we plan to incorporate migration policies into our empirical specification. Third, we want to disaggregate migration flows by type of immigration to account for the composition of migrants who participate in language courses and exams. Forth, we plan to digitize additional data from the records of the GI to extend our analysis to the period of 1980–2013 and to include information about language learning at institutes in Germany. This will also allow us to examine whether EU accession leads to a shift from pre-migration to post-migration language learning.

## References

Adsera, Alicia and Mariola Pytliková (2012). *The Role of Language in Shaping International Migration*. IZA Discussion Paper 6333. Bonn, Germany: Institute for the Study of Labor (IZA).

- Aparicio Fenoll, Ainhoa and Zoe Kuehn (2014). *Does Foreign Language Proficiency Foster Migration of Young Individuals within the European Union?* IZA Discussion Paper 8250. Bonn, Germany: Institute for the Study of Labor (IZA).
- Belot, Michèle and Sjef Ederveen (2012). “Cultural barriers in migration between OECD countries”. *Journal of Population Economics* 25(3), 1077–1105.
- Belot, Michèle and Timothy J. Hatton (2012). “Immigrant Selection in the OECD”. *The Scandinavian Journal of Economics* 114(4), 1105–1128.
- Bleakley, Hoyt and Aimee Chin (2004). “Language Skills and Earnings: Evidence from Childhood Immigrants”. *Review of Economics & Statistics* 86(2), 481–496.
- (2010). “Age at Arrival, English Proficiency, and Social Assimilation among US Immigrants”. *American Economic Journal: Applied Economics* 2(1), 165–92.
- Chiswick, Barry R. and Paul W. Miller (1995). “The Endogeneity between Language and Earnings: International Analyses”. *Journal of Labor Economics* 13(2), 246–288.
- (2007). “Computer usage, destination language proficiency and the earnings of natives and immigrants”. *Review of Economics of the Household* 5(2), 129–157.
- (2014). *International Migration and the Economics of Language*. IZA Discussion Paper 7880. Bonn, Germany: Institute for the Study of Labor (IZA).
- Dustmann, Christian and Francesca Fabbri (2003). “Language proficiency and labour market performance of immigrants in the UK”. *The Economic Journal* 113(489), 695–717.
- Dustmann, Christian and Arthur van Soest (2001). “Language Fluency and Earnings: Estimation with Misclassified Language Indicators”. *Review of Economics & Statistics* 83(4), 663–674.
- Espenshade, Thomas J. and Haishan Fu (1997). “An Analysis of English-Language Proficiency among U.S. Immigrants”. *American Sociological Review* 62(2), 288–305.
- Goethe-Institut e.V. (2013). *Jahrbuch 2012/2013*.
- Grogger, Jeffrey and Gordon H. Hanson (2011). “Income Maximization and the Selection and Sorting of International Migrants”. *Journal of Development Economics* 95(1), 42–57.
- Ispording, Ingo Eduard and Sebastian Otten (2013). “The Costs of Babylon—Linguistic Distance in Applied Economics”. *Review of International Economics* 21(2), 354–369.
- Lazear, Edward P. (1999). “Culture and language”. *Journal of Political Economy* 107(6), S95.
- Lohmann, Johannes (2011). “Do language barriers affect trade?” *Economics Letters* 110(2), 159–162.
- Mayda, Anna (2010). “International Migration: a Panel Data Analysis of the Determinants of Bilateral Flows”. *Journal of Population Economics* 23(4), 1249–1274.

- Melitz, Jacques and Farid Toubal (2012). *Native language, spoken language, translation and trade*. Discussion Paper 8994. London: Centre for Economic Policy Research.
- Ortega, Francesc and Giovanni Peri (2013). "The Effect of Income and Immigration Policies on International Migration". *Migration Studies* 1(1), 47–74.