CESifo WORKING PAPERS

6511
2017

June 2017

A Macro-level Analysis of Adult-age Language Learning

Silke Uebelmesser, Severin Weingarten

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editors: Clemens Fuest, Oliver Falck, Jasmin Gröschl
www.cesifo-group.org/wp

An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: www.CESifo-group.org/wp

CESifo *Working Paper No. 6511*
Category 4: Labour Markets

A Macro-level Analysis of Adult-age Language Learning

Abstract

This article investigates the macro-level drivers of adult-age language learning. We construct a new dataset that covers German language learning in 77 countries (including Germany) for 1992-2006. Fixed-effects regressions show that language learning in the EU is strongly associated with immigration. Instead, immigration by non-EU citizens in associated with language learning in Germany. Additionally, trade flows are strongly associated with language learning in non-EU high-income economies.

JEL-Codes: F220, J240, J610.

Keywords: language skills, language learning, migration, migrant networks, migration policy.

Silke Uebelmesser
University of Jena
Carl-Zeiss-Str. 3
Germany – 07743 Jena
silke.uebelmesser@uni-jena.de

Severin Weingarten
University of Jena
Carl-Zeiss-Str. 3
Germany – 07743 Jena
severin.weingarten@uni-jena.de

June 3, 2017
The authors gratefully acknowledge financial support by the German Science Foundation (DFG) (project UE 124/2-1).

1 Introduction

Language skills are a crucial prerequisite for communication across country borders and for the international mobility of people and goods. In the area of international migration choice, common languages, linguistic distance (Adserà and Pytliková 2015; Isphording and Otten 2013), and language education (Aparicio Fenoll and Kuehn 2016) have been shown to affect migration decisions. With respect to immigrant integration, language skills have a large positive effect on wages (Dustmann and Soest 2001; Dustmann and Soest 2002) and employment probabilities (Dustmann and Fabbri 2003) and on other measures of social integration (Aldashev et al. 2009). In the area of international trade, countries with a common language and a smaller linguistic distance are estimated to trade considerably more (Isphording and Otten 2013; Melitz and Toubal 2014).

The aim of this study is to investigate the determinants of adult-age language learning. Based on earlier research on migration choice, migrants' language skills and international trade, we hypothesise that language learning should be positively associated with overall migration, student migration, migrant stocks, and trade flows.

To the best of our knowledge, our study is the first to explore the determinants of adult-age language learning on the country-level. There is, however, a large number of studies in the migration field that explore the determinants of individual migrants' language skills. Chiswick and Miller (2015, section 4) offer an extensive review of the literature and divide the determinants of language skills into three categories, which they dub "the three E's": exposure summarises the environment in which migrants live and communicate, efficiency captures age at migration, level of education and similar characteristics that enhance individuals' abilities to learn, and, lastly, economic incentives cover a mix of internal and external factors such as planned duration of stay and expected earnings gains. Since most of "the three E's" vary on the level of the individual migrant, studies on the determinants of migrants' language skills use censuses or surveys to obtain micro-level data.

While these studies provide important insights into the relationship between individual characteristics and language skills, they have an important limitation: typically, they explain the self-reported language skills of the respondents at the time of the collection of the data. By doing so, they ignore the timing of language learning: Foreign language acquisition at early ages occurs primarily at school or as a consequence of parents' preferences. For adults on the contrary, the decision to learn a language is more likely made in light of a specific migration decision. This difference between early-age and adult-age learning has implications for the causal interpretation of the relationship between language skills and migration. Early-age learning is unlikely to be affected by migration decisions that are made in later life, but the opposite direction of causality is quite plausible: early-age learning builds up language skills and migrants may very well sort into the destination

countries where these skills are most useful. Studies that are based on measures of language skills at some point in time after migration cannot distinguish between early-age and adult-age learning. As a consequence, they cannot disentangle pre-migration language skills that caused migrants to sort into a particular destination and language learning that occurred as a consequence of the migration decision. For convenience, we will refer to these channels as the sorting channel and the incentive channel. From the point of view of the policy maker, an understanding of the incentive channel is highly relevant, because it allows the targeting of language courses at groups of immigrants who are more likely to lack necessary language skills.

Our study isolates the incentive channel by using participation in language courses and exams rather than individuals' language skills as the dependent variable. The necessary data were collected from the yearbooks of the Goethe-Institut, a German association that maintains cultural institutes and provides German language courses in many countries of the world (see Uebelmesser et al. 2017 for details). The dataset used for our estimations reports the number of language exam participants in 136 institutes located in 76 countries for the period from 1992 to 2006 and, separately, the number of language course participants of 157 nationalities in Germany for the same period. We argue that especially the exam participation data from the Goethe institutes is a reasonably good proxy for language learning in the wider populations of the countries where the institutes are located. The exams are widely recognised, do not require participation in a language course at a Goethe institute, and the institutes themselves are accessible to learners of all demographics.

Results from fixed-effects panel-data estimations indicate that language learning in EU countries is largely driven by student and non-student immigration to Germany, but that this is not the case for non-EU countries. Instead, immigration by non-EU citizens is positively associated with language learning in Germany. In addition, migrant stocks in (not migrant flows to) Germany are strongly positively associated with language learning in the countries' of origin of the immigrants. Trade flows are not significantly associated with language learning; with the exception of non-EU high-income economies where they are the only significant determinant of language learning in our study.

In addition to our fixed-effects estimations, which discard all between-country variation in the data, we use a random-effects model that separates within and between-country effects. Between-country differences show that German is studied a lot more in countries that send more students to Germany. This relationship may be the cause or the consequence of a stronger cultural link or it may be an indication that student migrants study German before or after their stay in Germany.

The paper is organised as follows. Section 2 outlines our hypotheses regarding the relationship between language learning and migration, as well as between language learning

and trade. Section 3 gives an overview of the datasets used in our analysis, in particular of the language-learning data. Section 4 describes our empirical set-up and section 5 discusses our results. Section 6 presents several robustness checks and extends the empirical set-up to a random-effects model. Section 7 concludes with a summary and policy recommendations.

2 Hypotheses

Our first set of hypotheses relates language learning to migration and the second set relates language learning to trade. Generally, the mechanisms described in this section and the resulting hypotheses apply to all countries, but they may differ in strength for EU and non-EU countries, because EU citizens enjoy unrestricted access to the German labour market. This issue is picked up in section 5 when we discuss the results of our estimations. Similarly, the described mechanisms and resulting hypotheses apply to both language learning abroad and language learning in Germany when no explicit distinction is made.

2.1 Migration

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 incentive to learn the language spoken in their host country. A positive association between immigration and course and exam participation would be in line with an incentive effect, where individuals decide to take language courses to improve their host-country language skills. If our analysis supports this hypothesis, it will complement the results of previous survey-based studies where the presence of migrant's language skills could either be attributed to an incentive effect or to a sorting effect (see section 1).

Hypothesis 1 Immigration from another country to Germany is positively associated with language learning by citizens of that country.

Students from other countries who come to Germany to study at German universities are a special subgroup of immigrants, but a similar incentive rationale applies to them. They may even be more motivated to learn the language than other groups of immigrants, because German law makes it easy for them to stay if they can find employment within a year after completing their studies.

Hypothesis 2 Immigration of students from another country to Germany is positively associated with language learning by citizens of that country.

Migrant Stocks

There are several reasons why not only current migration flows, but also the presence of migrants from a particular country in Germany should increase language learning by those who live in that country or who migrated from that country to Germany.

For language learning in Germany, migrants who arrived in previous years may still be taking language courses; either because they did not have the opportunity or motivation to do so earlier or because they are simply continuing their education.

For language learning abroad, a large migrant community in Germany may lead to increased interest in the German language by those who remained in the country of origin, but belong to the social circle of those who migrated to Germany. These individuals may simply be curious about the German culture, but the migration experience of others may also lead them to prepare their own (potential) migration to Germany.

Hypothesis 3 The presence of a large number of migrants from another country in Germany is positively associated with language learning by citizens of that country.

Migration and Minority Language Concentrations

Minority language concentrations[1] 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 live in minority language concentration. 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 and is, thus, at least partially driven by an incentive effect. 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 4 The presence of a larger number of migrants from another country in Germany weakens the association between migration and language learning by citizens of that country.

2.2 Trade

Trade relationships may also be an important driver of language learning decisions. 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.

Since international trade partners are located outside of Germany by definition, we expect larger trade flows to be associated with more language learning abroad, but not in Germany.

[1] Other authors use the terms "ethnic enclaves" (e.g. Danzer and Yaman 2013) or "migrant networks" (e.g. Bertoli and Fernández-Huertas Moraga 2015).

Additionally, since trade barriers and negotiation costs are presumably lower within the EU, we expect the association to be stronger for trade with non-EU countries.

Hypothesis 5 Larger trade flows between Germany and another country are associated with more language learning in that country; particularly if the country is not a member of the EU.

3 Data

To test the hypotheses outlined above, we need data on language learning activities on the one hand, and migration and trade-related variables on the other hand. We begin by introducing our dataset on language learning at the Goethe institutes and continue with other variables below.

3.1 Dependent Variables: The Goethe-Institut Dataset

The Goethe-Institut (GI) is a German association that promotes the study of the German language and culture abroad. Since its foundation in 1951, it has been maintaining institutes in 89 countries, at many of which locals can study the German language and obtain language certificates which are widely recognised. The GI is mainly funded by the German government and through course fees (Goethe-Institut e.V. 2015).

The GI reports key statistics about language learning at each individual institute in its yearbooks. The dataset from which our dependent variables are taken was created by digitizing these statistics.[2] It has two parts: The first part covers language learning abroad. It reports yearly observations of course and exam participation at Goethe institutes around the world. We use the exam participation variable from this part of the dataset, because it has two advantages over course participation: First, the exams cover fixed amounts of learning content (e.g. the A1/1 level in the Common European Framework of Reference for Languages, CEFR) and are therefore not affected by decisions of individual Goethe institutes to offer courses that cover more or less material, depending on their length. Second, exam participation is open to learners who did not participate in a language course at one of the Goethe institutes. As a consequence, it may act as a better proxy for the extent of language learning activities in the respective country. The second part of the dataset reports the yearly number of language course participants at the Goethe institutes in Germany, disaggregated by the nationality of the participants. For our analysis, we use the time period from 1992 to 2006 for which data from both parts of the datasets and migration data for a large number of countries are available. The 'Abroad' dataset covers a total of 136 institutes in 76 countries and the 'Germany' dataset covers 157 nationalities.

[2]The data were manually transferred to spreadsheets and checked for plausibility. Country and city names were then harmonised to facilitate matching them with country-level and city-level datasets from other sources. See Uebelmesser et al. (2017) for details.

While the hypotheses presented in the previous section address language learning in general, our data are limited to learners that come into contact with a Goethe institute. Naturally, there are a large number of other language learning opportunities, including universities, private language schools, and internet platforms. For language learning abroad, we alleviate this problem by using exam participation numbers. They cover a larger group of learners who may have learned the language elsewhere but have their skills certified at a Goethe institute. Still, the multitude of alternative learnings opportunities gives rise to a number of concerns regarding the self-selection of language learners into courses offered by the Goethe institutes that we will address below. 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. Historical price data on language courses are not available to the best of our knowledge. However, table 7 in the appendix contains current price data on comparable language courses offered by the Goethe institutes and by other institutes in six cities in different countries.[3] While the data are far from complete or representative, they do not indicate that the Goethe institutes are usually the most expensive provider in the market.[4]

Goethe institutes are usually located in capitals and other major cities. The lack of institutes in rural areas is likely to lead to an under-representation 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

[3]Data comes from the websites of the course providers. The websites of non-Goethe-institute providers were found by searching Google for "language learning" and the name of the respective city in the native language of the respective country.

[4]In conversations with us, several employees of the GI have stated that their language courses are priced to be self-financing and that government funding is used for non-language-course-related activities. Additionally, the price policies of the individual institutes take the prices of local competitors into account.

younger generation of language students, which is more familiar with using the internet in general. While this difference may lead to an over-representation 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 the period of observation from 1992 to 2006 used in our analysis. Consequently, we are confident that the age bias only has a small, if any, effect on our results.

3.2 Explanatory Variables

In estimations where the dependent variable is language learning abroad, the explanatory variables are characteristics of the countries where the respective institutes are located or characteristics of the relationships of these countries with Germany. In estimations where the dependent variable is language learning in Germany, we use the characteristics of the countries of origin of the language learners and these countries' relationships with Germany instead.

Migration. Data on migrant flows and stocks come from the German Federal Statistical Office (Destatis). Yearly immigration flows for our period of observation from 1992 to 2006 are available from the German "Wanderungsstatistik". It documents the number of citizens of each country that relocate their primary residence to Germany in a given year. Since residence registration is mandatory in Germany, we expect this measure to appropriately reflect legal immigration. Data on migrant stocks are taken from the Central Register of Foreign Nationals ("Ausländerzentralregister", AZR).

Student migration. Data on student migration come from Destatis as well. A direct measure of the number of immigrants, who migrate for the purpose of studying, is not available for Germany. Instead, we rely on aggregate university statistics ("Hochschulstatistik") and use the number of foreign students who are enrolled in their first semester at a Germany university as a proxy for student immigration flows.

Trade flows. We proxy the intensity of trade relationship by total trade revenues, imports plus exports, with each country of interest in a given year. Trade revenues are provided by Destatis.

Other control variables. Data on country and city populations come from the UN World Population Prospects and World Urbanization Prospects datasets respectively. As a measure of countries' gross domestic product (GDP) we use 'rgdpe' from the Penn World Table (v8.0).

Merging the datasets for our explained and explanatory variables leaves us with 136 institutes in 76 countries for our 'Abroad' estimation and 157 nationalities for our 'Germany' estimation. Tables 1 and 2 report descriptive statistics for both datasets.

Table 1: Descriptive Statistics 'Abroad' Estimation

Variable	Min.	Median	Mean	Max.	% Zeros
Language Exams (abroad)	0.00	76.00	176.45	6458.00	3.14
Immigration ($\times 10^3$)	0.00	4.33	10.42	152.73	1.63
Migrant Stock ($\times 10^3$)	0.51	26.88	124.04	2110.22	0.00
First-year Students ($\times 10^3$)	0.00	0.27	0.73	4.89	0.00
Trade Rev. with DE ($\times 10^6$)	0.02	5.99	22.77	147.11	0.00
GDP per capita ($\times 10^3$)	0.23	9.88	14.93	48.59	0.00
Population ($\times 10^6$)	2.29	40.77	130.15	1291.70	0.00
Population City ($\times 10^6$)	0.10	2.55	4.38	35.86	0.00

Note: All values are rounded to two decimal places.

Table 2: Descriptive Statistics 'Germany' Estimation

Variable	Min.	Median	Mean	Max.	% Zeros
Language Students (in DE)	0.00	27.00	140.15	2781.00	12.71
Immigration ($\times 10^3$)	0.00	0.58	3.88	152.73	7.56
Migrant Stock ($\times 10^3$)	0.01	3.91	40.70	2110.22	0.00
First-year Students ($\times 10^3$)	0.00	0.03	0.24	4.89	10.36
Trade Rev. with DE ($\times 10^6$)	0.00	0.30	6.52	147.11	0.00
GDP per capita ($\times 10^3$)	0.15	5.23	9.82	106.40	0.00
Population ($\times 10^6$)	0.06	8.19	37.23	1291.70	0.00

Note: All values are rounded to two decimal places.

4 Empirical Setup

We are interested in the relationships between language learning on the one hand and several migration-related and trade-related variables on the other hand, where language learning is proxied by two different variables: exam participation at institutes abroad and course participation at institutes in Germany.

For the 'Abroad' specification, we observe one exam participation number for each institute in each year. 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. We use institute-level rather than country-level estimations, because they allow us to exploit more of the variance in our explained variable. This approach also avoids the problem of abrupt changes in country-level aggregates of our dependent variable when institutes open and close. For the 'Germany' specification, the geographical structure is more straightforward. We observe one number of course participants in Germany for each nationality in each year and run our estimations on the country level.

We use OLS regressions with institute/country-fixed effects and year-fixed effects for all of our estimations. As a consequence of the fixed-effects approach, our coefficients are

only identified by variation within institutes/countries. For exam participation abroad, we estimate

$$P_{ijt} = \alpha + \beta\, x_{jt} + \gamma\, \text{Immig}_{jt} \times \text{MigStock}_{jt} + \delta\, \text{Pop}_{ijt} + \eta_t\, D_t + \eta_i\, D_i + u_{ijt} \qquad (1)$$

and for course participation in Germany

$$P_{jt} = \alpha + \beta\, x_{jt} + \gamma\, \text{Immig}_{jt} \times \text{MigStock}_{jt} + \eta_t\, D_t + \eta_j\, D_j + u_{jt} \qquad (2)$$

where the indices reflect the dimensions across which variables vary: institute i, country j, and time t. P represents exam or course participation and x_{jt} is a vector of our main explanatory and control variables: student and non-student migration flows, migrant stocks, trade revenues, country population size, and country GDP. γ captures the interaction between migration flows and stocks that is of interest for hypothesis 3 on enclave effects. Pop_{ijt} is the population of the city where each institute is located and D_i, D_j, and D_t are institute, country, and year dummies, respectively. In equation (1) institute-level dummies capture both country-fixed and institute-fixed effects. α is an intercept and u_{ijt} and u_{jt} are error terms. In both estimations, the errors are assumed to be clustered on the country level.[5]

All non-dummy variables in our regressions enter in logs so that coefficients can be interpreted as elasticities. The main advantage of this approach is that it allows variation from countries of different sizes and with completely different magnitudes of language learning, migration, and trade flows to drive the results of our model. An estimation in levels would suffer from considerable heteroskedasticity and the results would necessarily be driven by a small number of countries that send a large number of migrants to Germany or trade with Germany a lot. While immigration from or trade with these countries may be economically more relevant because of its magnitude, our main interest is in identifying the mechanisms that drive language learning more generally and, thus, in using variation from as many countries as possible to identify our coefficients. Additionally, an estimation in levels would require that we specify to which extent institutes in cities of different sizes are exposed to changes in our country-level explanatory variables. For example, an absolute change in immigration from France should, in absolute terms, have a larger effect on exam participation in Paris than on exam participation in Nancy. Paris is a larger city and the institute there has a larger catchment area. A log-log estimation does away with this concern, because it assumes that both institutes experience the same relative rather than absolute change in exam participation as a result of a relative increase in migration.[6]

[5] Cluster-robust standard errors are calculated according to Cameron et al. (2011) using the R package multiwayvcov (version 1.2.2).

[6] A small proportion of the values of our dependent and independent variables are zero. For a log-log estimation we have to eliminate these values. To do so, we replaced all zero values with 0.5 before applying

5 Results

In this section, we present the results of our fixed-effect estimations and discuss to which extent these results can be interpreted as evidence of causal effects.

5.1 Main Estimation Results

Our 'Abroad' estimations use a total of 1,464 observations from 136 institutes in 76 countries. The dependent variable is the number of language exams taken at institute i in year t. Our 'Germany' estimations use a total of 2,261 observations for 157 nationalities. In these estimations, the dependent variable is the number of language courses taken by individuals with nationality j in year t. All regressions cover the time period from 1992 to 2006 and are run separately for EU countries and for non-EU countries. This leads to a total of four regressions. Table 3 reports the results.

With respect to hypotheses 1 and 2, we find a significant positive association between language learning on the one hand and both general immigration and student immigration on the other hand. While this is not the case in all four estimations, the pattern of the results is quite clear: Migration from EU countries is associated with language learning abroad, while migration from non-EU countries is associated with language learning in Germany.

A potential explanation for this pattern is that EU citizens can move to and work in Germany without any restrictions. They can safely invest in pre-migration language learning, because their investments are not exposed to the risk of becoming obsolete if entry to Germany is denied. Non-EU citizens, on the contrary, are exposed to this risk and may therefore defer language learning until after their arrival in Germany. Vice versa, language learning in non-EU countries is not associated with migration flows that occur in the same year and EU citizens do not seem to participate in language courses at the Goethe institutes after their arrival in Germany. In the latter case at least for student migration, however, it seems likely that the receiving universities provide (further) training in the German language.

The sizes of the coefficients suggest that the association between language learning and migration is somewhat stronger for language learning in EU countries than for the language learning of non-EU citizens in Germany. Taking general immigration and student migration together, an increase in immigration by 1% leads to an increase in language learning in

the log. This strategy seems sensible for two reasons: First, all affected variables are positive integers so that 0.5, like 0, is smaller than all other values. Second, the distributions of all affected variables have some mass close to zero. Replacing 0 with 0.5 keeps the distinction between original zeros and original ones intact and keeps the logs of the original zeros relatively close to the logs of the other small but non-zero observations.

Table 3: Estimation results for course and exam participation abroad and in Germany

	Abroad		Germany	
	EU	non-EU	EU	non-EU
Immigration	0.54**	−0.06	0.10	0.04***
	(0.27)	(0.04)	(0.18)	(0.01)
Foreign First-year Students	0.01	0.08	0.37	0.18***
	(0.30)	(0.11)	(0.25)	(0.04)
Migrant Stocks	0.52	0.67**	0.19	0.11
	(0.91)	(0.29)	(0.42)	(0.10)
Imm. × Mig. Stock	−0.00	−0.01	−0.07	−0.01
	(0.13)	(0.03)	(0.08)	(0.01)
Trade Rev. w/ Ger.	0.31	0.63**	0.47	0.05
	(0.81)	(0.30)	(0.43)	(0.06)
GDP per capita	−1.89	0.25	−0.48	0.33*
	(1.80)	(0.36)	(0.82)	(0.17)
Population	10.29	−1.87	1.32	−0.56
	(6.47)	(3.32)	(3.24)	(0.58)
Population City	5.27	2.14		
	(3.59)	(1.61)		
R^2	0.70	0.65	0.98	0.91
Adj. R^2	0.65	0.61	0.97	0.90
Num. obs.	369	1225	224	2026
Num. institutes	35	112	–	–
Num. countries	16	67	23	145
Num. years	15	15	15	15

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Standard errors are clustered on the country level. Observations from countries that joined the EU during the period of observation are assigned to the EU and non-EU samples on the basis of the year of ascension to the EU.

EU countries by 0.76%, but only to an increase in language learning by non-EU citizens in Germany of 0.41%.

With respect to hypothesis 3, we find a strong positive association between migrant stocks in non-EU countries and the language learning that occurs there. This association cannot be driven by the preparatory language learning of those who immigrated in the same year, because this effect would be captured by the coefficient for immigration flows. Similarly, it cannot be the consequence of a causal link between immigration and trade (Bandyopadhyay et al. 2008; Peri and Requena-Silvente 2010), because all of our specifications include trade flows as a separate explanatory variable. Instead it points towards the mechanism described in section 2, where larger migrant communities in Germany lead to an increased interest in the German language abroad.[7]

[7]One might also suspect that migrant stocks act as a proxy for short-term migration to Germany that is not captured by our residence-based measure of migration flows. However, while this may be the case, it seems unlikely that short-term migrants would have a sufficiently large incentive to learn German, to

With respect to hypothesis 4, we find no clear indication that the association between language learning and migration flows is moderated by the size of migrant stocks. The coefficient for the interaction between migration flows and migrant stocks is barely significant in the estimation for language learning in non-EU countries, but the main effect of migration itself is insignificant in this estimation. We conclude that our data does not support the conjecture that language learning efforts by migrants decrease as diasporas in Germany grow.

With respect to hypothesis 5, we do not find a significant association between language learning and trade.

5.2 Direction of Causality

Above, we provide evidence of a positive association between language learning and several migration-related and trade-related variables. In this section, we argue that this association is mainly driven by a causal effect of those variables on language learning.

With respect to language learning in Germany, a reverse causal effect of course participation on our migration-related variables is unlikely, because language learners are already located in Germany by definition. Theoretically, the availability of post-migration language courses might affect the decision whether and where to migrate. However, since courses that teach migrants the language of the destination country are likely to be available in any potential destination, the potential effect of availability on migration can be neglected in our application.

With respect to language learning at institutes outside of Germany, three channels of reverse causality could be particularly relevant: the opening and closing of institutes, the participant recruitment and capacity planning of the institutes, and changes in the individual motivation of participants. First, the opening and closing of institutes may be motivated by the presence of migration and trade flows. This channel of causality would affect the results of our estimations if they "compared" participation in cities in which language courses or exams were offered with zero participation in cities where this was not the case. However, our dataset only includes observations for city-year combinations where language courses were offered and is, thus, not susceptible to the endogenous opening and closing of institutes.[8] Second, the presence of large migration flows from a particular country may motivate institutes in that country to advertise more heavily to "capture" a larger share of the outgoing migrants. While we cannot measure the intensity of advertising of individual institutes, there is no indication that the institutes follow such a strategy. In our conversations with officials at the Goethe institutes, they repeatedly

drive the strong association between migrant stocks and language learning reported above.

[8] See Huber and Uebelmesser (2017) for an analysis the effect of language learning opportunities via the opening of institutes on migration to Germany.

stated that they attempt to adjust to local demand rather than to actively encourage outgoing migrants to participate in courses. Third, the language learning experience at the Goethe institutes may motivate individuals to move to Germany, who initially take the course for non-migration related reasons. While this may be a good description of the experience of a small number of language learners, the migration choice literature seems to agree that the key determinants of migration decisions are others, income differentials and migration policies in particular (Grogger and Hanson 2011; Ortega and Peri 2013; Bertoli and Fernández-Huertas Moraga 2015).

With respect to trade, language learning helps to overcome linguistic barriers, but it is unlikely to have a sizeable short-run effect on trade volumes. Such an effect would imply that a large proportion of individuals begin language courses for non-trade related reasons, but then decide to use their newly acquired skills to establish trade relationships with Germany. While this may happen in individual cases it is unlikely to happen often, because trade is usually conducted through firms and has considerable set-up costs.

6 Robustness Checks and Extensions

In the following, several robustness checks are presented and the empirical set-up is extended to a random-effects model.

6.1 Capitals and Large Cities

One potential concern with our 'Abroad' estimation is that institutes in cities with different characteristics attract language learners from different sub-groups of the population of the respective country and that our estimations throw all of these sub-groups together. In particular, institutes in capital cities may attract a large number of language learners that work for the government or for large businesses. While we cannot control for the individual characteristics of different language learners in our macro-level study, table 4 presents the results of four estimations for language learning abroad, where we restrict the sample to institutes from capital and non-capital cities.

With respect to the positive association between language learning and general migration from EU countries, the results do not point towards a difference between capitals and non-capitals. Somewhat unsurprisingly, the immigration coefficient is no longer significant, because standard errors increase as we split the already small EU sample in half. However, the coefficient has roughly the same size in both estimations.

The picture changes as we shift our attention towards student immigration. Its positive association with language learning in EU countries seems to be largely driven by non-capital cities.

Table 4: Estimation Results: Capitals and Large Cities

	EU		non-EU	
	Cap	Big	Cap	Big
Immigration	0.33**	0.47***	−0.10*	−0.04
	(0.13)	(0.16)	(0.05)	(0.03)
Foreign First-year Students	0.03	0.24	0.24	0.15
	(0.25)	(0.25)	(0.15)	(0.15)
Migrant Stocks	0.63	0.49	0.75**	0.74**
	(1.19)	(0.98)	(0.30)	(0.30)
Imm. × Mig. Stock	0.02	−0.11	−0.05	−0.00
	(0.19)	(0.20)	(0.04)	(0.02)
Trade Rev. w/ Ger.	0.04	−0.02	0.83**	0.70**
	(0.59)	(0.55)	(0.39)	(0.34)
GDP per capita	−2.89	−2.63	0.35	0.28
	(2.21)	(2.22)	(0.53)	(0.35)
Population	8.87	13.26	−0.52	−0.30
	(10.11)	(9.22)	(3.65)	(3.46)
Population City	5.90**	8.73***	2.82	2.12
	(2.42)	(2.27)	(1.96)	(1.87)
R^2	0.91	0.91	0.70	0.66
Adj. R^2	0.88	0.88	0.66	0.62
Num. obs.	165	165	706	810
Num. institutes	16	16	60	67
Num. countries	16	16	59	67
Num. years	15	15	15	15

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Standard errors are clustered on the country level. Observations from countries that joined the EU during the period of observation are assigned to the EU and non-EU samples on the basis of the year of ascension to the EU. The "Cap" sample is smaller than the "Big" sample, because 8 countries do not have an institute in their capital.

The opposite is true for the association between language learning and migrant stocks found for non-EU countries, a result that is entirely contained to capital cities. This suggests that migrant stocks lead to increased language learning by employees of the government or of large businesses whose headquarters are located in capital cities.

6.2 High-income vs Lower-income Countries

In addition to heterogeneity across different types of cities, one can argue that language learning has different determinants in different countries. The relative importance of the motives for language learning may depend on additional characteristics of the countries that are not captured by our control variables. In this section, we explore differences between countries with different income levels.

Migration-related motives may be more relevant to those who live in low-income countries, because their potential income gains from migrating to Germany are higher. Vice-versa, trade-related motives may be more important in high-income countries: More complex and more technology-intensive trade relationships with other highly developed countries

Table 5: Estimation Results: High-income vs Low-income Countries

	non-EU	High Inc. non-EU	Lower Inc. non-EU
Immigration	−0.24	−0.07	−0.25*
	(0.15)	(0.80)	(0.15)
Foreign First-year Students	0.14	0.34	0.13
	(0.11)	(0.33)	(0.12)
Migrant Stocks	0.69***	−0.61	0.77***
	(0.24)	(0.58)	(0.26)
Imm. × Mig. Stock	−0.11*	−0.09	−0.13*
	(0.07)	(0.24)	(0.07)
Trade Rev. w/ Ger.	0.26	1.11***	0.18
	(0.19)	(0.41)	(0.21)
GDP per capita	0.30	−0.65	0.36
	(0.26)	(1.90)	(0.27)
Population	−2.01	−0.33	−2.39
	(1.86)	(1.58)	(2.19)
Population City	1.68	−2.81*	2.65*
	(1.28)	(1.54)	(1.36)
R^2	0.75	0.84	0.74
Adj. R^2	0.72	0.81	0.70
Num. obs.	1125	287	838
Num. institutes	111	31	80
Num. countries	67	14	53
Num. years	15	15	15

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Standard errors are clustered on the country level. Observations from countries that joined the EU during the period of observation are assigned to the EU and non-EU samples on the basis of the year of ascension to the EU.

may come with higher language requirements. Management or R&D careers in German-speaking international companies are more likely in highly industrialized countries, where these kind of jobs are typically located.

The World Bank provides a classification of economies by income-level. We use the classification from 2006, the last year covered by our dataset, to split our sample into high-income countries and countries with a lower ranking (lower-income countries). We omit EU countries, because virtually all EU countries were high-income economies in 2006 and the remaining lower-income sample would be too small to conduct a meaningful analysis.[9]

Table 5 presents the results of our estimations. The coefficient for trade flows is large and significant in the high-income sample; suggesting that a 1% increase in trade volume with Germany translates into a 1.1% increase in language learning at institutes in the respective country. The migration-related coefficients are not significantly different from zero in this

[9]The only two non-high-income countries in our EU sample are Latvia and Poland. They joined the EU in 2004 and our dataset ends in 2006. This makes for a total of nine observations: three observations each for the institutes in Riga, Krakow, and Warsaw.

sample. Vice-versa, the coefficient for trade flows is not significantly different from zero in the lower-income sample, but the coefficient for migrant stocks is. A 1% increase in migrant stocks translates into a 0.8% increase in language learning.

These results highlight a stark difference between the role of language learning in countries with different income levels. Business relationships are a crucial motive for language learning in high-income countries, while migration-related motives are dominant in lower-income countries.

6.3 Within-Between Estimations

While our fixed-effects approach eliminates potential biases from time-invariant city-specific and country-specific factors that are correlated with language learning and our explanatory variables, it also prevents us from exploring between-country differences. However, these differences may be interesting in their own right. Within-country changes in our variables are often smaller than the between-country differences and a pure within-country analysis could therefore miss important parts of the "bigger picture". Naturally, between-country comparisons have to be interpreted very carefully, because they are not robust to the omission of time-invariant variables that are correlated with both the left-hand side and the right-hand side of the estimation equation.

Between-country differences may also be interesting because the exact timing of the relationship between language learning and our explanatory variables is unclear. For example, potential migrants may learn the language several years before migrating to Germany or they may do so several years after their arrival. We could try to capture these types of lagged associations with a large series of lags, however, our 15-year-long dataset is too short to support such an approach. Additionally, the individual lagged effects may be too small to obtain statistically significant coefficients and the timing of language learning may differ across countries.

To explore between-country differences, we complement our fixed-effects models with a random-effects (RE) specification. We use Bell and Jones' (2014) re-formulation of an RE estimator proposed by Mundlak (1978). Instead of removing all between country variance from the data, the RE estimator uses it to identify the between coefficients (subscript B). As with the FE estimator, the within coefficients (subscript W) are only identified by

within-country variation. For exam participation abroad, we estimate

$$
\begin{aligned}
P_{ijt} = {} & \alpha + \beta_W \left(x_{jt} - \bar{x}_j\right) + \beta_B \, \bar{x}_j \\
& + \gamma_{WW} (\text{Immig}_{jt} - \overline{\text{Immig}}_j) \times (\text{MigStock}_{jt} - \overline{\text{MigStock}}_j) \\
& + \gamma_{WB} \, \overline{\text{Immig}}_j \qquad\qquad\quad \times (\text{MigStock}_{jt} - \overline{\text{MigStock}}_j) \\
& + \gamma_{BW} \, (\text{Immig}_{jt} - \overline{\text{Immig}}_j) \times \overline{\text{MigStock}}_j \\
& + \gamma_{BB} \;\; \overline{\text{Immig}}_j \qquad\qquad\quad \times \overline{\text{MigStock}}_j \\
& + \delta_W \left(\text{Pop}_{ijt} - \overline{\text{Pop}}_i\right) + \delta_B \, \overline{\text{Pop}}_{ij} \\
& + u_i + u_j + u_t + u_{ijt}
\end{aligned}
\tag{3}
$$

and for course participation in Germany

$$
\begin{aligned}
P_{jt} = {} & \alpha + \beta_W \left(x_{jt} - \bar{x}_j\right) + \beta_B \, \bar{x}_j \\
& + \gamma_{WW} (\text{Immig}_{jt} - \overline{\text{Immig}}_j) \times (\text{MigStock}_{jt} - \overline{\text{MigStock}}_j) \\
& + \gamma_{WB} \, \overline{\text{Immig}}_j \qquad\qquad\quad \times (\text{MigStock}_{jt} - \overline{\text{MigStock}}_j) \\
& + \gamma_{BW} \, (\text{Immig}_{jt} - \overline{\text{Immig}}_j) \times \overline{\text{MigStock}}_j \\
& + \gamma_{BB} \;\; \overline{\text{Immig}}_j \qquad\qquad\quad \times \overline{\text{MigStock}}_j \\
& + u_j + u_t + u_{jt}
\end{aligned}
\tag{4}
$$

where bars indicate averages. This within-between specification also disentangles the interaction between immigration and migrant stocks into three components: One component that is driven by the within variation of both variables (WW) and two additional components, each of which is driven by the within variation of one but by the between variation of the other variable (WB and BW). For example, the within-between interaction "Immigration (B) × Migrant Stocks (W)" can be interpreted as the change in the effect of the country-average of immigration on language learning that is brought about by growing (or shrinking) migrant stocks. A fourth interaction term is identified by the between variation of both variables (BB), but this term has no equivalent in the FE estimation, where it would have been captured by the institute-fixed or country-fixed effects. u_i, u_j, u_t, u_{jt} and u_{ijt} are error terms (random effects). These specifications are extensions of Bell and Jones' (2014) equation (12). We add the interaction terms and the additional error term u_i to account for our nested geographical structure of institutes within countries.

We would expect most of the coefficients that are identified by within variation to be the same in our RE and FE estimations, because the regression models are very similar. The two differences are the disentangled interaction term discussed above and the different error structure. In addition to the FE models' idiosyncratic error term, u_{ijt} or u_{jt}, the RE models allow for random errors on the levels of institutes, countries, and years. The

expectation of identical within coefficients is confirmed by the results presented in table 6. The between model adds an interesting results.

In addition to the significant within coefficient of student migrant stocks for German institutes discussed above, the between coefficient of student migrant stocks is large and significant in both the 'Abroad' and the 'Germany' specification for non-EU countries. This implies that countries that, averaged across the entire period of observation, sent more students to Germany also showed more language learning activity, both abroad and in Germany. There are at least three potential explanations for this result. First, in the long term, student migration may lead to increased interest in the German language and culture abroad. Second, strong cultural links may not be the consequence but the cause of both language learning activities and migration. While time-invariant cultural links are controlled out of our within coefficients, they do enter into the between coefficients. Third, student migrants may learn German several years before or after studying in Germany. As discussed above, this effect would not be picked up by our unlagged within coefficient, but may show up in the between coefficient.

7 Conclusions

In this paper, we use records of the German Goethe-Institut to construct a new dataset on the extent of German language learning around the world. It covers course and exam participation at Goethe institutes in Germany and 76 other countries for the period from 1992 to 2006. To the best of our knowledge this is the first large-scale dataset on adult-age language learning. We use the exams variable from the 'Abroad' dataset and the participant number from the 'Germany' dataset for our estimations. Both measures vary considerably between and within institutes and countries and this variance is not explained by differences in population alone.

We use the dataset to investigate the determinants of language course and exam participation abroad and in Germany. Our results complement those of studies which use individual-level datasets to investigate the determinants of language skills, but not actual learning decisions, of migrants. We add to the literature by disentangling language learning decisions that are made in the context of migration and other economic factors from the presence of language skills that may have been either the cause or the result of migration decisions. We argue that our results can be interpreted in terms of incentives to learn German.

Using fixed-effects regressions, we find that language learning at Goethe institutes in the EU is strongly associated with immigration, but that this is not the case for non-EU countries. Instead, the immigration of non-EU citizens is positively associated with language learning in Germany. This suggests that preparatory language learning is much

more likely for immigration from the EU than from other countries. Language learning at institutes in non-EU countries is instead strongly associated with migrant stocks in Germany.

The lack of a positive association between language learning in non-EU countries and migration from these countries to Germany should be taken into account by policy makers. It supports the notion that many non-EU migrants arrive in Germany without language skills and that policy interventions may be necessary to foster preparatory language learning. Since EU citizens do seem to learn German in preparation of their migration, attention should be payed to differences between the incentives of EU and non-EU immigrants. For example, non-EU immigrants are faced with considerable uncertainty as to whether and how long they will be allowed to stay in the country. In turn, this uncertainty may decrease the incentive to invest in language skills prior to the resolution of this uncertainty.

Table 6: Estimation Results: Within-between Random-effects Specification

	Abroad		Germany	
	EU	Non-EU	EU	Non-EU
Immigration (W)	0.57**	−0.04	0.15	0.04***
	(0.23)	(0.04)	(0.11)	(0.01)
Immigration (B)	−1.02*	0.01	0.59	0.13*
	(0.57)	(0.23)	(0.52)	(0.08)
Foreign First-year Students (W)	−0.05	0.00	0.32**	0.16***
	(0.28)	(0.07)	(0.15)	(0.03)
Foreign First-year Students (B)	1.36**	0.67***	−0.19	0.41***
	(0.58)	(0.20)	(0.41)	(0.05)
Migrant Stocks (W)	−0.26	0.62***	−0.02	0.12*
	(0.93)	(0.15)	(0.44)	(0.07)
Migrant Stocks (B)	1.25***	−0.24	−0.82**	−0.06
	(0.47)	(0.22)	(0.39)	(0.08)
Imm. (W) × Mig. Stock (W)	−2.18	0.32*	0.06	0.04
	(2.15)	(0.18)	(0.83)	(0.06)
Imm. (W) × Mig. Stock (B)	0.27*	−0.00	−0.05	−0.01
	(0.15)	(0.02)	(0.07)	(0.01)
Imm. (B) × Mig. Stock (W)	−0.27	−0.12	0.04	−0.03
	(0.77)	(0.08)	(0.32)	(0.03)
Imm. (B) × Mig. Stock (B)	−0.39**	−0.04	−0.04	−0.03***
	(0.17)	(0.04)	(0.05)	(0.01)
Trade Rev. w/ Ger. (W)	−0.00	0.36***	0.52**	0.05
	(0.62)	(0.12)	(0.25)	(0.04)
Trade Rev. w/ Ger. (B)	1.47***	0.17	−1.36***	0.12**
	(0.48)	(0.21)	(0.44)	(0.06)
GDP per capita (W)	−2.55**	0.21	−1.07*	0.30***
	(1.11)	(0.21)	(0.54)	(0.08)
GDP per capita (B)	−3.77**	−0.47*	2.35**	0.43***
	(1.64)	(0.27)	(1.19)	(0.08)
Population (W)	5.20	−4.25***	−0.68	−1.00***
	(4.83)	(0.92)	(2.35)	(0.30)
Population (B)	−2.62***	−0.37**	2.66***	0.32***
	(0.69)	(0.16)	(0.49)	(0.05)
Population City (W)	3.38	2.48***		
	(2.56)	(0.65)		
Population City (B)	0.56***	0.35**		
	(0.18)	(0.14)		
Num. obs.	369	1225	224	2026
Num. institiutes	35	112		
Num. countries	16	67	23	145
Num. years	15	15	15	15
Var: Institute-level	0.64	1.01		
Var: Country-level	0.19	0.33	0.71	0.28
Var: Year-level	0.00	0.00	0.01	0.02
Var: Residual	0.63	0.90	0.10	0.45

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Standard errors were calculated from a bootstrap ($n = 10{,}000$); significance levels are based on basic confidence intervals calculated from the same bootstrap sample. Observations from countries that joined the EU during the period of observation are assigned to the EU and non-EU samples on the basis of the year of ascension to the EU.

A Language Course Pricing

City	Provider	Course Type	Price / Hour	Currency
Mexico City	GI	Extensive	146.67	MXN
Mexico City	Tecnologico 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 7: Prices of language courses at Goethe institutes and other providers in 2015

References

Adserà, Alícia and Mariola Pytliková (2015). "The Role of Language in Shaping International Migration". *The Economic Journal* 125(586), F49–F81.

Aldashev, Alisher, Johannes Gernandt, and Stephan L. Thomsen (2009). "Language usage, participation, employment and earnings: Evidence for foreigners in West Germany with multiple sources of selection". *Labour Economics* 16(3), 330–341.

Aparicio Fenoll, Ainhoa and Zoë Kuehn (2016). "Does Foreign Language Proficiency Foster Migration of Young Individuals within the European Union?" In: *The Economics of Language Policy*. Ed. by Bengt-Arne Wickstroem and Michele Gazzola. CESifo Seminar Series. Cambridge, MA: MIT Press.

Bandyopadhyay, Subhayu, Cletus C. Coughlin, and Howard J. Wall (2008). "Ethnic Networks and US Exports". *Review of International Economics* 16(1), 199–213.

Bell, Andrew and Kelvyn Jones (2014). "Explaining Fixed Effects: Random Effects Modeling of Time-Series Cross-Sectional and Panel Data". *Political Science Research and Methods* FirstView, 1–21.

Bertoli, Simone and Jesús Fernández-Huertas Moraga (2015). "The size of the cliff at the border". *Regional Science and Urban Economics* 51, 1–6.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller (2011). "Robust Inference With Multiway Clustering". *Journal of Business & Economic Statistics* 29(2), 238–249.

Chiswick, Barry R. and Paul W. Miller (2007). "Computer usage, destination language proficiency and the earnings of natives and immigrants". *Review of Economics of the Household* 5(2), 129–157.

Chiswick, Barry R. and Paul W. Miller (2015). "International Migration and the Economics of Language". In: *Handbook of the Economics of International Migration*. Ed. by Barry R. Chiswick and Paul W. Miller. Vol. 1A. Handbooks in Economics. Oxford: North Holland, 211–269.

Danzer, Alexander M. and Firat Yaman (2013). "Do Ethnic Enclaves Impede Immigrants' Integration? Evidence from a Quasi-experimental Social-interaction Approach". *Review of International Economics* 21(2), 311–325.

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.

— (2002). "Language and the Earnings of Immigrants". *Industrial & Labor Relations Review* 55(3), 473–492.

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. (2015). *Jahrbuch 2014/2015*.

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.

Huber, Matthias and Silke Uebelmesser (2017). "Presence of Language-learning Opportunities abroad and Migration to Germany". *University of Jena, mimeo*.

Isphording, 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.

Melitz, Jacques and Farid Toubal (2014). "Native language, spoken language, translation and trade". *Journal of International Economics* 93(2), 351–363.

Mundlak, Yair (1978). "On the Pooling of Time Series and Cross Section Data". *Econometrica* 46(1), 69–85.

Ortega, Francesc and Giovanni Peri (2013). "The Effect of Income and Immigration Policies on International Migration". *Migration Studies* 1(1), 47–74.

Peri, Giovanni and Francisco Requena-Silvente (2010). "The trade creation effect of immigrants: evidence from the remarkable case of Spain". *Canadian Journal of Economics/Revue canadienne d'économique* 43(4), 1433–1459.

Uebelmesser, Silke, Matthias Huber, and Severin Weingarten (2017). "The German Language Worldwide: A New Dataset on Adult-Age Language Learning". *University of Jena, mimeo.*