



INSTITUTE FOR EMPLOYMENT  
RESEARCH  
The Research Institute of the Federal Employment Agency

# IAB-DISCUSSION PAPER

Articles on labour market issues

---

**14|2019** Say it like Goethe: Language learning facilities  
abroad and the self-selection of immigrants

Philipp Jaschke, Sekou Keita



ISSN 2195-2663

# Say it like Goethe: Language learning facilities abroad and the self-selection of immigrants

Philipp Jaschke (IAB)

Sekou Keita (IAB)

Mit der Reihe „IAB-Discussion Paper“ will das Forschungsinstitut der Bundesagentur für Arbeit den Dialog mit der externen Wissenschaft intensivieren. Durch die rasche Verbreitung von Forschungsergebnissen über das Internet soll noch vor Drucklegung Kritik angeregt und Qualität gesichert werden.

The “IAB Discussion Paper” is published by the research institute of the German Federal Employment Agency in order to intensify the dialogue with the scientific community. The prompt publication of the latest research results via the internet intends to stimulate criticism and to ensure research quality at an early stage before printing.

## Contents

Abstract . . . . .	4
Zusammenfassung . . . . .	4
1 Introduction . . . . .	6
2 Institutional background of Goethe Institutes . . . . .	10
3 Theoretical background . . . . .	11
4 Data . . . . .	14
5 Empirical analysis . . . . .	16
5.1 Descriptive statistics . . . . .	20
5.2 Main results . . . . .	25
5.3 Effects on other characteristics . . . . .	25
5.4 Robustness checks and sensitivity analyses . . . . .	26
5.4.1 Goethe Institutes without language services . . . . .	26
5.4.2 Demand for German language classes . . . . .	28
5.4.3 Heterogeneity of main results . . . . .	28
6 Mediation analysis . . . . .	30
6.1 Mechanism . . . . .	31
6.2 Transmission channels . . . . .	33
7 Conclusions . . . . .	36
References . . . . .	37

## Abstract

Immigration policy in most high-income countries is designed to promote qualified migration while maintaining high requirements on characteristics such as education and language skills. We rely on a standard self-selection model with heterogeneous migration costs to discuss the effect of access to language learning services in the country of origin on the skill composition of immigrants in Germany. Using individual-level survey data on immigrants from different cohorts over the period 2000–2014, combined with unique data on the presence of Goethe Institutes — a German association promoting German language and culture worldwide — in origin countries, the results of our empirical analysis show that the acquisition of the German language is fostered by the availability of language courses abroad. Moreover, we find that language services abroad induce a positive (self-)selection of migrants along several dimensions, such as education, experience, and the probability of holding a job offer at arrival. These characteristics are in turn highly relevant for long-term integration in Germany. To disentangle transmission channels, we perform a causal mediation analysis. We find that 25 % of the total effect of language services abroad on language skills at immigration trace back directly to migrants' participation in language courses, revealing important spillover effects.

## Zusammenfassung

Die Einwanderungspolitik in den meisten einkommensstarken Ländern zielt darauf ab, qualifizierte Migration bei gleichzeitiger Beibehaltung hoher Anforderungen an Bildung und Sprachkenntnisse zu fördern. Wir verwenden ein etabliertes theoretisches Modell zur Selbstselektion von Migrantinnen und Migranten mit heterogenen Migrationskosten, um die Auswirkungen des Zugangs zu Sprachlernangeboten im Herkunftsland auf verschiedene Qualifikationen von Migrantinnen und Migranten in Deutschland zu untersuchen. Ausgehend von individuellen Befragungsdaten von Zugewanderten aus Kohorten zwischen 2000 und 2014, kombiniert mit neu vorliegenden Daten über die Präsenz von Goethe Instituten in den Herkunftsländern, zeigen unsere Ergebnisse, dass der Erwerb der deutschen Sprache durch die Verfügbarkeit von Sprachkursen im Ausland gefördert wird. Darüber hinaus stellen wir fest, dass Sprachlernangebote im Ausland eine positive (Selbst-)Selektion von Migrantinnen und Migranten entlang mehrerer Dimensionen bewirken, wie z.B. Bildung, Erwerbserfahrung und die Wahrscheinlichkeit eines Stellenangebots bei Einreise. Diese Eigenschaften sind wiederum für die langfristige Integration in Deutschland von großer Bedeutung. Zur Identifikation von Wirkungskanälen, führen wir eine kausale Mediationsanalyse durch. Es zeigt sich, dass 25 % des gesamten Effekts durch angebotene Sprachkurse im Ausland auf die Sprachkenntnisse zum Zeitpunkt der Einwanderung direkt auf die Teilnahme von Migrantinnen und Migranten an Sprachkursen zurückzuführen sind. Dies deutet auf substantielle Spillover-Effekte hin.

**JEL classification:** F22, J18, J24, J61, Z13

**Keywords:** International Migration, Public Policy, Human Capital – Skills, Immigrant Workers

**Acknowledgements:** We are grateful to the participants in the IAB internal seminars and the IAB and CERGE-EI workshop on “Labor market challenges in times of globalization, technological, and demographic change” as well as the participants in Masaryk University’s “Young Economists’ Meeting” for excellent comments.

[...] It is “astonishing how astonishing it still is that they are coming”, says Holger Kolb, at the Expert Council of German Foundations on Integration and Migration. Some things are beginning to work as intended, such as the elimination of bureaucratic hassles for moving within the EU. Yet, it seems that the EU can never become a truly integrated market. That is mainly because of language. Mr Gómez finds Germans challenging—“always nagging you about recycling or noise or whatever”—but the language is “the hardest part”. Thus language has replaced work visas as the main barrier to mobility. ”

---

The Economist, *Sprechen Sie job?* , February 16, 2013.

[...] The government also focuses on the promotion of German language skills. For example, the Goethe Institutes abroad are to be given greater support, vocational German courses are to be expanded, and vocational training programs with integrated language training are to be developed.<sup>a</sup>

---

<sup>a</sup> Authors' own translation. The original text is the following: “Einen Schwerpunkt legt die Regierung auch auf die Förderung von deutschen Sprachkenntnissen. So, sollen etwa die Goethe-Institute im Ausland stärker gefördert, berufsbezogene Deutschkurse ausgebaut und Berufsbildungsprogramme mit integrierter Sprachausbildung entwickelt werden. ”

Handelsblatt, *Das sind die 5 Kernpunkte des geplanten Einwanderungsgesetzes*, August 16, 2018.

## 1 Introduction

The optimal design of immigration policy is one of the most controversial topics in both policy debates and the academic literature. On the one hand, wealthy countries generally prefer to welcome highly skilled migrants with good integration potential while limiting the arrival of less educated migrants. On the other hand, potential migrants from poorer countries face high barriers that prevent them from moving to countries where the rewards for their labor are potentially higher. In both perspectives, language skills play an important role. Since the seminal contributions by Chiswick (1978, 1991), many studies have shown that knowledge of the language of the destination country is one of the most important aspects of economic and social integration. As a consequence, language requirements are part of the procedure to obtain a work visa in many countries. However, learning a foreign language from abroad is a far-reaching investment, both financially and in terms of effort. Such high costs could prevent qualified candidates from exploiting migration opportunities. This is especially true for destination countries such as Germany with languages that are less widespread in the world. As noted in the article in *The Economist* cited above, language acts as an important barrier to mobility even in a world where administrative hurdles to mobility have in many cases been removed. The German government seems to have placed this issue higher on its agenda since active support for German learning facilities abroad is explicitly part of the discussion around the new immigration law which was approved in the parliament in June 2019. However, insights from the literature on the

determinants of language learning decisions in the home countries of potential migrants are scarce. The selection effects of such language learning opportunities are even less well understood.

The aim of this paper is to identify the effect of the presence of German language learning facilities in the country of origin on immigrants' language skills at arrival. We adapt a standard self-selection model with heterogeneous migration costs to discuss the effect of increased access to language learning services in the country of origin on the skill composition of migrants.

Next, we conduct an empirical analysis on a sample of 2,486 immigrants originating from 84 countries who arrived in Germany between 2000 and 2014. The analysis sample is obtained by merging individual-level information from the IAB-SOEP Migration Sample with data on the number of Goethe Institutes with language services in immigrants' countries of origin taken from Uebelmesser/Huber/Weingarten (2018).<sup>1</sup> Drawing on variation in the number of institutes present in the country of origin one year prior to migration, across immigrant cohorts and within countries of origin, our estimates reveal that the effect of exposure to German language learning facilities on several characteristics which are closely linked to improved labor market integration. In particular, our results show that exposure to German language learning facilities significantly improves the self-reported language skills at arrival. We also find that the number of language institutes present in the country of origin prior to migration improves selection in terms of education and labor market experience. Consistent with the language requirements of German immigration law, immigrant cohorts with greater exposure to language learning institutes are more likely to hold a work visa at arrival.

Although our identification strategy considerably reduces potential bias from omitted variables, one might nevertheless be concerned about the influence of time-varying factors that correlate with migrant cohort quality and the opening or closing of language service facilities in a given country of origin. We show that this threat to identification is not supported by the data. First, none of the major determinants of migrant selection significantly explain variations in the number of Goethe Institutes. Second, we show that our results do not hold when we consider variation in Goethe Institutes without language services. Third, the presence of language services after migration does not influence the characteristics of migrants at arrival, suggesting that the variation in the number of active institutes is not driven by demand in the origin country. Finally, we rely on a formal mediation analysis to disentangle different transmission channels. Specifically, we quantify the share of the total effect mediated through participation in language courses abroad, isolating the effect induced by other selection patterns. We find that 25 percent of the total effect of exposure to German language learning institutes on language skills at immigration is mediated through actual participation in language classes abroad. This suggests that such language learning facilities have important spillover effects and contribute to overcoming migration barriers through several channels.

---

<sup>1</sup> In the following, the term "Goethe Institute" refers to those institutes that provide language learning services if not otherwise stated.

This paper is related to two major strands of the literature. First, we address the literature on the self-selection of immigrants. Tracing back to the contribution by Borjas (1987), who builds on Roy (1951), the question of ‘who migrates?’ has received considerable attention. At the core of most theoretical models are income or wage differences. The results in Ramos (1992) and Borjas (2008) on migration between Puerto Rico and the US are in line with predictions from the Borjas (1987) theoretical model with constant migration costs. The model states that migrants from countries with low wage inequality between low- and high-skilled workers to countries with high wage inequality should be positively selected. This result is challenged by evidence from Chiquiar/Hanson (2005), who find moderately positive selection of migrants between Mexico and the United States. These results can be rationalized in a model with heterogeneous migration costs with respect to worker skills. However, a subsequent study by Fernández-Huertas Moraga (2011) reports negative selection in the same migration corridor after addressing data limitations. Grogger/Hanson (2011) suggest that absolute wage level differences within skill groups between countries play a more central role. They argue that a linear utility model fits the empirical results on selection in Mexican-US migration better than the Borjas model because earnings equality in Mexico is higher than in the US (Hanson, 2006), which would imply negative selection.

Nevertheless, reliable empirical evidence based on micro data is relatively scarce considering all the variables that could shape self-selection. For instance, migrant networks constitute an important variable that affects self-selection because the availability of friends and/or relatives in the destination country can decrease migration costs (Munshi, 2003; McKenzie/Rapoport, 2010). Bertoli (2010) identifies migrants’ networks as a driver of negative self-selection with respect to the education of Ecuadorian emigrants that moved to a set of destination countries. McKenzie/Rapoport (2010) document similar evidence for emigrants from Mexico to the US.

To the best of our knowledge, Bauer et al. (2002) provide the only evidence for self-selection of immigrants to Germany by matching cross-sectional micro data for the years 1982 and 1986 on Portuguese workers who stayed in Portugal and those who emigrated to Germany. On the one hand, they document an overall negative selection of emigrants from Portugal working in Germany with respect to observable educational characteristics, a finding that is likely related to the guest worker contracts that were agreed with Portugal and through which the majority of individuals in the sample came to Germany. On the other hand, when comparing wage outcomes between immigrant Portuguese workers with their German counterparts within the group of blue collar workers, they find higher wages for the Portuguese, suggesting a positive selection with respect to unobserved characteristics.

Evidence from previous studies leaves little room for policy to improve the self-selection of immigrants, as the most studied determinants, such as income differentials, income inequality, and networks, cannot easily be influenced by policy in the short run. Some studies have investigated the success of various immigration policies in attracting highly skilled immigrants, e.g., the point-based systems adopted in Canada, Australia and New Zealand, which allow for substitution between required characteristics, as opposed to the restrictive system in place in Germany, which requires migrants to simultaneously satisfy all requirements (for an overview, see Brücker et al. (2012)). However, such selection

criteria “at the border” mostly have an effect after the self-selection of migrants in their origin country has already occurred. As Bauer et al. (2002: pp. 467) state, “[...] any immigration policy can only choose from the pool of people willing to migrate”.

Our paper contributes to this literature by investigating the selection effect of a reduction in migration costs induced by the presence of language learning facilities abroad, a variable on which policy can act. The results suggest that this is an effective way to increase the qualification level of future immigrant cohorts.

Second, our paper contributes to the literature investigating the role of language skills in the performance of migrants at the destination. Language proficiency has been consistently identified as crucial for the integration of immigrants in the host country’s society and labor market. One channel identified in prior research suggests that language enhances the transferability of human capital from the origin country (Dustmann/Glitz, 2011). Additionally, post-migration investments in human capital, such as schooling, are more effective if migrants are proficient in the host country’s language (Van den Berg, 1994; Chiswick/Miller, 2003; Zschirnt/Ruedin, 2016). Moreover, language proficiency determines wages and explains significant parts of wage convergence between immigrants and natives (Chiswick, 1998). Nevertheless, heterogeneity could exist between high- and low-skill occupations (Berman/Lang/Siniver, 2003). To properly identify the effect of language skills on wages, it is important to account for omitted factors such as unobserved ability that biases the estimates of language ‘directly’ (Borjas, 1994), through self-selection into occupations (Berman/Lang/Siniver, 2003) or educational institutions (Bleakley/Chin, 2004).

Multiple approaches have been applied in the literature to overcome endogeneity and identify the causal effect of language on various outcomes. Instruments used include the ethnic concentration in the father’s birthplace, having married abroad and the existence and age of children (Chiswick/Miller, 1995), father’s education (Dustmann/van Soest, 2001; Dustmann/Soest, 2002), immigrants’ age at arrival (Bleakley/Chin, 2004), being parent of a language-proficient child and willingness to stay in the host country in the coming years (Budria/Swedberg, 2015), and language skills at immigration (Schuss, 2018). These studies have provided evidence underlining the importance of language for successful integration. Our paper contributes to this literature by highlighting the role of language learning opportunities abroad in shaping language skills at immigration. Furthermore, we show that improved language skills coincide with modified selection patterns in many other characteristics associated with productivity. This result calls for caution when researchers interpret the estimated effect of language on labor market performance.

The remainder of this paper is organized as follows: Section 2 presents the institutional framework of the Goethe Institutes. Section 3 provides a theoretical framework for the immigrant selection effect expected from language learning facilities, which is followed by a description of the data in Section 4. Section 5 presents the empirical strategy and results of the analysis, including robustness checks and sensitivity analyses. Transmission channels are discussed in greater detail in Section 6 using a formal mediation analysis. Finally, Section 7 provides some concluding remarks.

## 2 Institutional background of Goethe Institutes

The Goethe Institute was founded in 1951 and exclusively entrusted with the promotion of German as a foreign language. From the 1960s onwards, the range of tasks was extended such that it increasingly developed into a cultural organization. At present, the Goethe Institute is one of the main actors in German foreign cultural policy and operates in many parts of the world as a central information and contact point for language courses, German culture, lifestyle and politics.

The Goethe Institute is a non-profit organization that pursues exclusively and directly charitable purposes (Goethe Institute, 2018b). The main source of funding is the German government. Nevertheless, the association retains content-related autonomy ensured by a treaty with the Federal Foreign Office (Goethe Institute, 2016). The current version of the treaty identifies three main duties of the Goethe Institutes. The first is the promotion of the German language through courses, training for teachers, and scholarships for language learning. The second is support for international cultural cooperation and involves cultural actors in the destination countries in cultural events and exchanges. The third is the provision of information about social, political, and cultural life in Germany by making available libraries with German books and media and by organizing cultural exchange with Germany (Uebelmesser/Huber/Weingarten, 2018).

According to official figures, the organization is currently represented by 159 institutes in almost 100 countries (Goethe Institute, 2018a). More than three-quarters of the facilities offer language courses. In addition, there are extensive educational partnerships with tens of thousands of regular schools abroad that offer German lessons. Note that language courses provided by the Goethe Institute are not subsidized by the German government. Instead, they are financed solely by course fees.

The Goethe Institute has a right of proposal; however, decisions on openings and closings take place in joint consultation between the German Federal Foreign Office and the Goethe Institute. The decision process involves discussions between the two stakeholders on the basis of their general objectives, considering the legal, political, and social aspects of the host country. Beyond the Federal Foreign Office, the influence of other political representatives and institutions is very limited. Importantly, representatives from either business or the Federal Ministry of Economics have no say or decision-making rights at any point. The Presidium of the Goethe Association consists of 12 members, one of whom is delegated by the Federal Foreign Office and another delegated by the Federal Ministry of Finance. The Presidium is elected by the General Assembly, which consists primarily of (i) representatives of various branches of the cultural, scientific and social life of the Federal Republic of Germany (full members) and (ii) the group of so-called “extraordinary members”. The latter comprises a representative of the parliamentary groups of parties sitting in the German parliament (Bundestag) and two representatives of the German state governments. For more detailed information regarding the institution’s organization and political independence, we refer to a report provided by the scientific service of the German Bundestag (Wissenschaftliche Dienste des Deutschen Bundestages, 2006).

With respect to the empirical strategy of our analysis, the crucial aspect of the institutional

background of the Goethe Institute is that, during the period of analysis, its actions were not driven by concerns related to the employment or migration policy of the German government.

### 3 Theoretical background

To guide our empirical analysis, we rely on a standard self-selection model of international migrants with heterogeneous migration costs that decrease with education. Borjas (1987) was the first to adapt the model of self-selection developed by Roy (1951) to the issue of international migration using constant migration costs. Borjas (1991) further developed the model by introducing heterogeneous migration costs. Building on this extension, Chiquiar/Hanson (2005) show that if migration costs are decreasing in skill, migration incentives are such that positive, negative, and intermediate selection all become plausible depending on the level of migration costs and the shape of the skill distribution in the country of origin. Further sources of heterogeneity, such as migration networks, can be added to migration costs to analytically assess their implications for self-selection (McKenzie/Rapoport, 2010).

We adopt the same notation as Chiquiar/Hanson (2005) and McKenzie/Rapoport (2010) to write the wage equations in the country of origin and the country of destination as

$$\ln w_0 = \mu_0 + \delta_0 s \quad (1)$$

$$\ln w_1 = \mu_1 + \delta_1 s \quad (2)$$

where  $w$  is the wage,  $\mu > 0$  is the minimum wage when schooling is 0, and  $\delta > 0$  stands for the returns to schooling, which is represented by  $s$ . The subscripts 0 and 1 stand for the origin country and the destination country, respectively. In line with the literature, we assume that minimum wages are systematically higher in the country of destination ( $\mu_1 > \mu_0$ ) and returns to schooling in the country of origin exceed those in the country of destination ( $\delta_0 > \delta_1$ ).

Migration costs, denoted  $C$ , can be expressed in time-equivalent units such as

$$\pi = \frac{C}{w_0}. \quad (3)$$

The migration decision can then be expressed as a comparison between wages in the destination with wages in the country of origin, net of migration costs. Specifically, migration occurs if <sup>2</sup>

<sup>2</sup> We use the following approximation:  $\ln(1 + \pi) \simeq \pi$ . The validity of the approximation requires  $\pi$  to be sufficiently small. This appears reasonable given that  $w$  is defined as the present value of future wages (McKenzie/Rapoport, 2010).

$$\ln w_1 - \ln (w_0 + C) \cong \ln w_1 - \ln w_0 - \pi > 0 \quad (4)$$

where time-equivalent migration costs  $\pi$  depend on a constant term  $\mu_\pi$ , an education-specific component  $\gamma_1 > 0$ , and a component capturing the effect of the presence of language services offered by Goethe Institutes  $\gamma_2 GI$  with  $\gamma_2 > 0$  and  $GI \geq 0$  such that

$$\ln \pi = \mu_\pi - \gamma_1 s(1 + \gamma_2 GI) \iff \pi = e^{\mu_\pi - \gamma_1 s(1 + \gamma_2 GI)}. \quad (5)$$

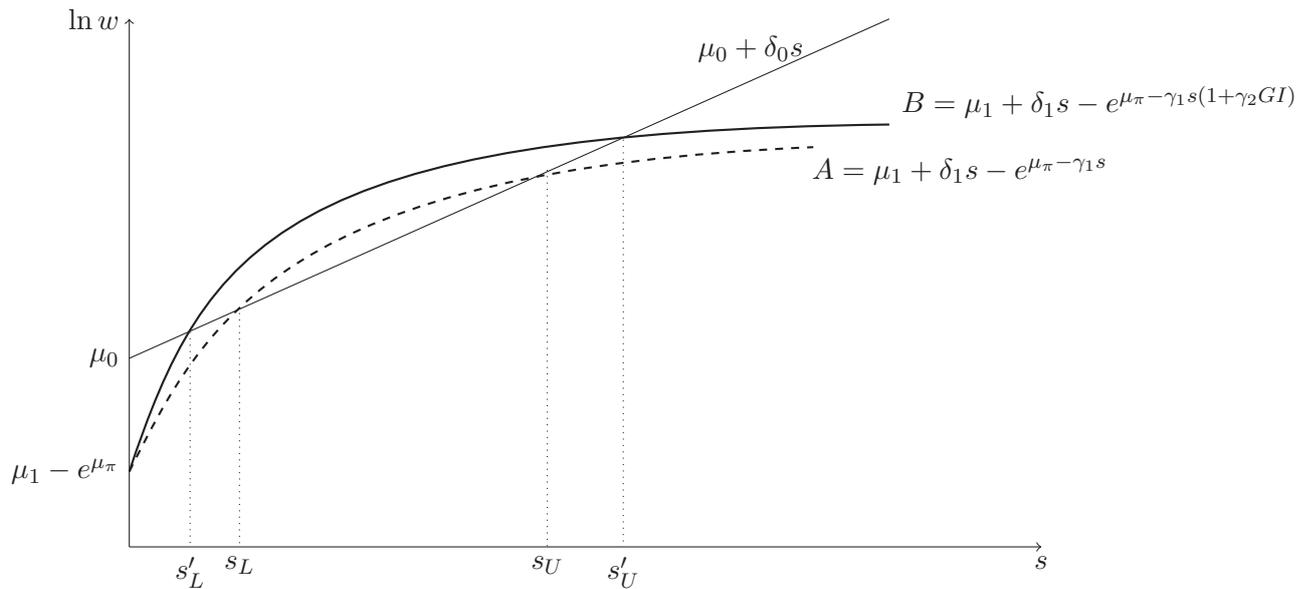
Equation (5) clarifies the assumptions regarding time-equivalent migration costs. In line with Chiquiar/Hanson (2005) and McKenzie/Rapoport (2010), migration costs decrease with schooling for several reasons. First, to the extent that higher education increases wage rates, better educated workers will require fewer hours of work to pay fixed migration costs. Second, better educated workers are more efficient at coping with intangible migration costs such as administrative paperwork and processing vast amounts of information when migration is legal or bargaining with smugglers when migration is illegal.

The number of language service facilities offered by Goethe Institutes  $GI$  in the country of origin also reduces migration costs. This is mainly because living and working in Germany requires investing in a minimum level of language skills to successfully navigate recurring transactions such as administrative procedures, housing contracts, transportation and commuting, media and information consumption, and communication with coworkers and civil servants. Acquiring language skills implies forgone earnings because workers cannot work as many hours (or not work at all) during a significant portion of their language training period, which can vary from several weeks to several years. As both living costs and wages are higher in Germany than in most countries of origin, the possibility to learn the language before migrating therefore represents a significant reduction in migration costs. We assume that the cost-reducing effect of language learning services described above is heterogeneous with respect to education. Specifically, we assume that increased language services are a complement to education. For workers without education, the presence of language service facilities does not affect migration costs. The rationale is that taking advantage of formal language classes requires basic skills that workers without education lack, such as reading and writing. Once a minimum level of education is available, every additional unit of education reinforces the reduction in time-equivalent migration costs induced by the presence of language service facilities. This occurs for several reasons. First, language learning efficiency is likely to increase with education up to a certain level, implying that better educated workers require less time to achieve a given level of language proficiency. Second, because the language services offered by Goethe Institutes are expensive relative to income levels in most countries of origin, better educated workers earning higher wage rates are more likely to be able to afford the classes. Third, a positive correlation between education and earnings also implies that the opportunity cost of learning German is higher for better educated workers. Therefore, they would benefit disproportionately from the increased possibility of learning the language in the home country. Fourth, Goethe Institutes can also reduce migration costs through other channels beyond the provision of language classes by providing information about the German society and labor market, generating

network effects through German expatriates, and facilitating access to media and documentation in German. The ability to exploit such additional effects is also likely to increase with the education level of workers.

In addition, we assume that  $e^{\mu\pi} > \mu_1 - \mu_0$ , i.e., migration costs for workers with no education are high enough to offset the migration incentives induced by the differential in minimum wages between the two countries. This assumption ensures that we do not exclude the possibility of positive selection (Chiquiar/Hanson, 2005; McKenzie/Rapoport, 2010).

Figure 1: Presence of Language Learning Services and Self-Selection of Migrants



The main insights from this theoretical framework are illustrated in Figure 1. In the absence of Goethe Institutes, the wage profile faced by potential migrants is given by the dashed curve  $A = \mu_1 + \delta_1 s - e^{\mu\pi - \gamma_1 s}$ . Workers migrate if their expected wage at the destination exceeds their wage in the country of origin, i.e., the dashed curve is above the line  $\mu_0 + \delta_0 s$ . Without language services, the incentives are such that migrants self-select from a schooling interval between  $s_L$  and  $s_U$ . Below  $s_L$ , migration costs are too high, and migration is not profitable. Above  $s_U$ , the returns to education in the country of origin are high enough to offset any migration incentive.

The effect of Goethe Institutes is to increase migration incentives for all schooling levels above 0, shifting the wage profile upwards as depicted by the solid curve  $B = \mu_1 + \delta_1 s - e^{\mu\pi - \gamma_1 s(1 + \gamma_2 GI)}$  in Figure 1. It is straightforward that more people will migrate following a reduction in migration costs.<sup>3</sup> Stated formally, we have the following:

<sup>3</sup> Note that this result is consistent with the findings of Huber/Übelmesser (2019), who show that a greater presence of language services abroad through Goethe Institutes raises migration flows towards Germany and German-speaking regions in Switzerland.

## Proposition 1

*Language learning services abroad increase overall migration incentives for workers with a strictly positive level of education.*

PROOF: See the appendix.

Compared to the situation with no Goethe Institutes present in the country of origin, migration incentives are increased for workers at both extremes of the schooling distribution. The new thresholds are such that  $s'_L < s_L$  and  $s'_U > s_U$ . The implications of introducing or expanding Goethe Institutes abroad for self-selection depend on the sign of  $|s_L - s'_L| - |s_U - s'_U|$ . Because the effect of language services is a complement to education, the effect of Goethe Institutes on migration incentives should be stronger for higher levels of education. This can be shown formally.

## Proposition 2

*Language learning services abroad increase migration incentives more for workers with a high level of education.*

PROOF: See the appendix.

As noted by Borjas (2014), the selection patterns predicted by the model are highly sensitive to the assumptions on how a particular variable affects migration costs and returns to migration. Hence, the effect of introducing or expanding Goethe Institutes abroad on the self-selection of migrants is ultimately an empirical question.

## 4 Data

Our analysis builds on a sample combining variables from four different sources. Specifically, we match (i) individual-level variables on immigrants in Germany with (ii) the origin-specific presence of language learning facilities prior to migration, (iii) origin-specific labor market conditions in Germany at the time of immigration, and (iv) economic conditions in the country of origin prior to migration.

### IAB-SOEP Migration Sample

The individual-level variables included in the analysis draw on the IAB-SOEP Migration Sample (Brücker et al., 2014), a dataset produced by a collaboration between the Institute for Employment Research (IAB) and the Socio-Economic Panel (SOEP) at the German Institute for Economic Research (DIW). The first wave of the survey was launched in 2013, including 4,964 respondents residing in 2,723 households. Follow-up interviews based

on this initial sample were then conducted in 2014, 2015, and 2016. In 2015, a refresher sample of 1,096 households containing 1,689 adult respondents who migrated to Germany between 2009 and 2013 was added to the survey.

An important feature of the IAB-SOEP Migration Sample is that it includes biographical surveys covering education and labor market histories of respondents in their country of origin. Furthermore, the survey also reports retrospective information on skills, administrative status, and reasons for migration at the time of immigration. This combination of information on characteristics both before migration and at the time of immigration is essential for our analysis because we are interested in the effect of pre-migration exposure to German language learning facilities on selection with respect to variables determining integration potential.

The sample used for our analysis is restricted to adults between 18 and 65 years old who immigrated to Germany less than 16 years prior to the interview date.<sup>4</sup> Further sample restrictions are imposed by the availability of information on the main variables. Our main sample for analysis includes 2,486 individuals; a detailed description of the sample with respect to the main variables used in the analysis is presented in Section 5.1.

## **Presence of language learning facilities**

Data on the presence of Goethe Institutes – as the most important German association promoting German culture and language worldwide – are obtained from Uebelmesser/Huber/Weingarten (2018). The authors kindly made publicly available a unique dataset on the presence of Goethe Institutes offering language courses, standardized exams and cultural services in more than 100 countries over the period 1965–2015. The main source of information is the annual reports of the Goethe Institutes.

The dataset created by Uebelmesser/Huber/Weingarten (2018) provides annual information on the number of Goethe Institutes in various cities in each country. They are usually present in the largest cities, implying that access to German learning facilities is more costly for people living in rural areas. Nevertheless, the schedule of language classes is relatively flexible, offering intensive courses that allow participants from different regions to take advantage of the classes.

The data allow us to disentangle the presence of Goethe Institutes with and without language services. In our analysis, we use the total number of Goethe Institutes offering language services in each country-year observation as a measure of exposure to German language learning. Hence, variation for different cohorts originating from a specific country comes from the number of institutes opening or closing in the year prior to migration.<sup>5</sup>

---

<sup>4</sup> The rationale for this restriction is to reduce measurement error in the self-reported language skills at immigration. The prior is that the likelihood of misreporting language skills at immigration increases with time elapsed between the immigration year and the interview year due, for example, to recall bias.

<sup>5</sup> For some countries, the exact number of Goethe Institutes providing language services is unavailable because numbers are reported jointly for more than one Goethe Institute in a single annual report. Moreover, note that changes in our main explanatory variable over time reflect either the opening (closing) of an entire Goethe Institute, the beginning (termination) of providing language services, or both.

Details on the variability in the number of Goethe Institutes over time and across countries of origin are presented in Section 5.1.

## Macro-level variables

We employ data on labor market conditions in Germany in the year of immigration as control variables. The statistics department of the German Federal Employment Agency (BA) provides annual data on the employment rate (employees divided by population, both aged between 15 and 64) as well as the mean and median wage disaggregated by nationality and year, which we merge to individuals based on origin countries and immigration years (German Federal Employment Agency, 2018). Note that the wage data are censored from above at the social security contribution limit for each marginal euro earned (approximately 5,800 € monthly wage in east and 6,500 € in west Germany). Therefore, we include the median wage, which safeguards our estimates against this source of bias. The statistical population for this information consists of employees who are compulsorily registered for health, pension and statutory nursing care insurance. Also included are trainees, interns and part-time employees. Civil servants, self-employed persons, family workers, and soldiers and people in military or alternative service are excluded. Employees with multiple employments are only included once with their most recent position.

For origin countries, we control for the unemployment rate and log GDP per capita (2011 USD, PPP) one year before immigration. The data stem from the World Development Indicators (Worldbank, 2018). Data on linguistic proximity of origin countries relative to Germany stem from Melitz/Toubal (2014) and are available for download on CEPII's website. Specifically, we use the "common language index based on level specification".

## 5 Empirical analysis

The aim of our empirical analysis is to identify the effect of the presence of German language learning facilities on the selection of migrants. As previously discussed, because the "quality" of immigrants is a multidimensional concept, we consider the effect of exposure to German language learning facilities on several characteristics that in turn determine the pace of integration in Germany. Specifically, we estimate the following equation:

$$Y_{iot} = \alpha_1 + \beta_2 GI_{o(t-1)} + \beta_3 X_i + \beta_3 Z_{o(t-1)} + \delta_o + \delta_t + \varepsilon_{iot} \quad (6)$$

where  $Y_{iot}$  is an individual-level outcome variable measuring characteristics for immigrant  $i$  originating from country  $o$  who arrived in Germany in year  $t$ . We focus on characteristics defined at the time of immigration or prior to immigration that are relevant for subsequent integration and labor market performance. In particular, our main outcome variables are language skills at immigration, level of education attained in the country of origin, pre-migration labor market experience, and status at immigration. Note that there are concerns about measurement problems when asking about (especially retrospective) language skills

in surveys (Dustmann/van Soest, 2001). Bleakley/Chin (2004) provide evidence mitigating this concern by revealing high correlations between language proficiency reported in surveys and objective test scores. In addition, measurement errors due to potential recall bias in the dependent variables would only be a concern for the results of our analysis if they systematically coincided with the exposure to language facilities prior to migration.

The variable  $GI_{o(t-1)}$  is our main variable of interest and measures the number of language facilities present in country of origin  $o$  one year before migration.  $X_i$  is a set of time-invariant socioeconomic variables measured at the time of immigration, including gender, age, work experience prior to immigration (in years), and the level of education attained in the country of origin.  $Z_{o(t-1)}$  is a set of macroeconomic variables capturing the economic conditions both in the country of origin one year prior to migration and in Germany. Controlling for economic conditions in the country of origin is crucial because they are likely to capture important push factors and opportunity costs that determine the size and composition of migration flows. Similarly, economic conditions in Germany represent important pull factors that also shape the composition of migration flows. If labor market conditions in Germany or in the country of origin are correlated with variations in the number of language facilities present in the country of origin and shape the self-selection patterns of migrants with respect to characteristics such as language proficiency, for example, our estimated coefficient  $\beta_2$  would be biased due to omitted variables. We address this concern by using GDP per capita and the unemployment rate as proxies for economic conditions in the country of origin. For each country of origin, we also use median wages in Germany computed for workers holding the corresponding nationality for each immigration year  $t$ . Importantly, all regressions include a set of country fixed effects  $\delta_o$  controlling for all country-specific, time-invariant determinants of the outcome variable under consideration. Such time-invariant determinants include, but are not limited to, linguistic and cultural distance, quality of education, the education system, and unobserved preferences.<sup>6</sup> All regressions also include a full set of immigration-year fixed effects  $\delta_t$ , which account for time-varying factors specific to Germany affecting immigrants from all countries of origin. Such factors include, among others, economic conditions in Germany, overall German migration policy, and electoral cycles. The term  $\alpha_1$  refers to the intercept, while  $\varepsilon_{1iot}$  indicates an idiosyncratic error term.

The coefficient of interest in our analysis  $\beta_2$  captures the effect of the number of language learning facilities present in country of origin  $o$  one year before migration on the outcome variable. The inclusion of country fixed effects  $\delta_o$  ensures that only variation within a given country of origin is used for identification. In other words, our estimated effect is obtained by comparing German language skills for immigrant cohorts originating from the same country who have been exposed to different numbers of language facilities present in their country of origin one year before migration. One crucial assumption for the identification of  $\beta_2$  is the absence of correlation between  $GI_{o(t-1)}$  and  $\varepsilon_{1iot}$  conditional on all other control variables. This condition would be violated if, for example, the opening of language learning facilities systematically coincided with the enforcement of a labor mobility agreement

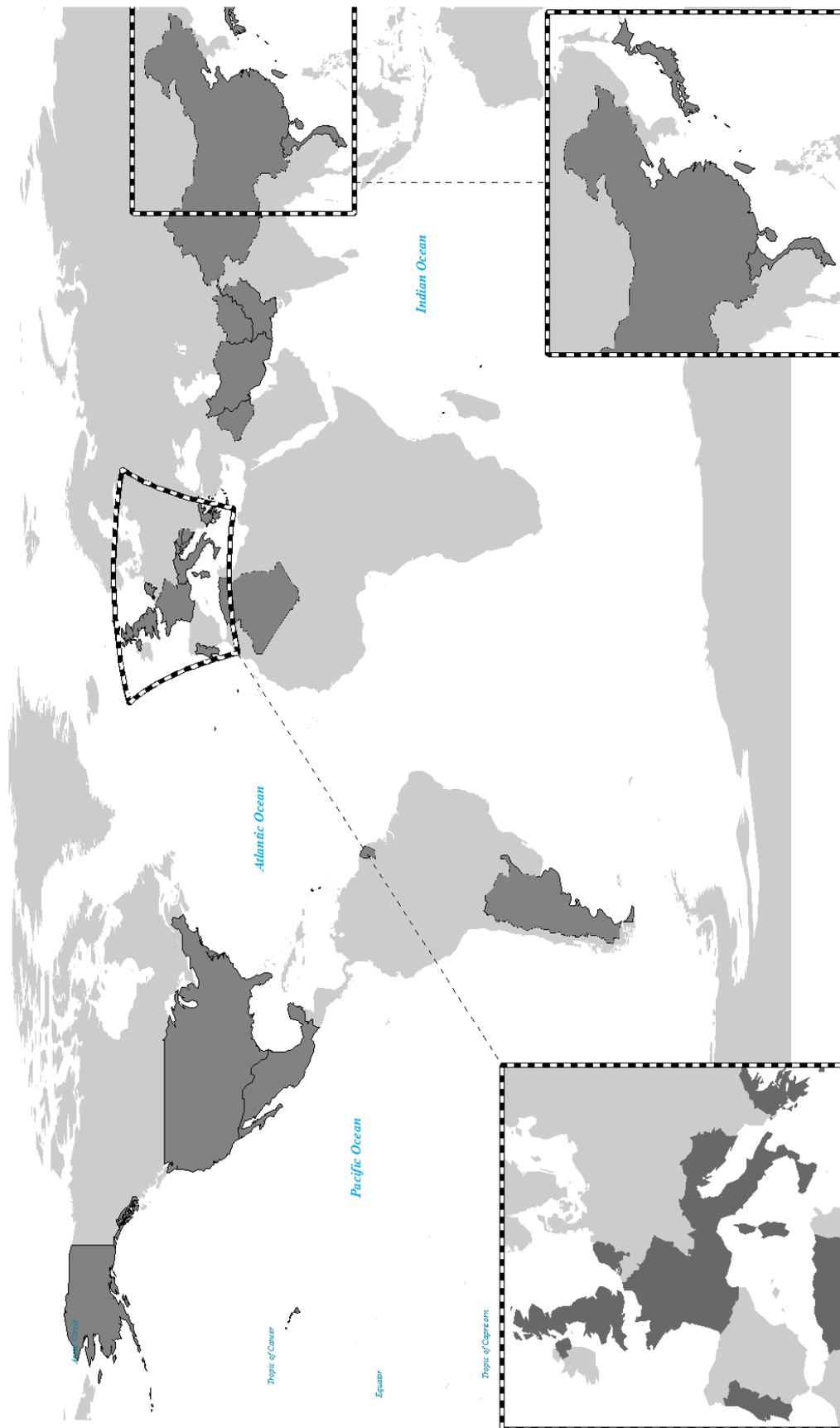
---

<sup>6</sup> While the overall quality of the education system as a whole undoubtedly contributes to determining the ability to learn any foreign language, one can also imagine differences directly related to the German language. For example, the national education program in some countries offers the opportunity to learn German as a second language in high school or even before.

lowering the barriers to immigration between Germany and the country of origin. In other words, the validity of our identification strategy requires the expansion plans of the Goethe Institute to be unrelated to the labor market or migration policy of the German government. Otherwise, our results would capture the effect of unobserved skill-selective migration policies. Moreover, our identification strategy requires that the variation in the presence of language learning opportunities abroad is not driven by demand for German classes due to the intention to move to Germany. We first discuss these threats to identification from an institutional perspective, focusing on the aims pursued by the Goethe Institute and the decision-making process for openings and closings. Second, we also address the potential issues empirically by investigating the determinants of openings and closings of language learning facilities abroad and by providing robustness checks of our main results in the form of placebo regressions.

The institutional setting of the Goethe Institute, presented in Section 2, indicates that actions of the Goethe Institute organization are not systematically related to the labor market or migration policy of the German government. Figure 2 further supports this argument. It maps the countries in our working sample that are driving our empirical identification, namely those presenting variation in the number of Goethe Institutes over time and from which we observe at least two individuals that were exposed to a different number of Goethe Institutes one year prior to migration. As can be seen, identification rests on variation across all continents, with a focus on Western Europe and Asia. If the expansion of Goethe Institutes were systematically linked to migration policy, one would expect a focus on countries with a large diaspora in Germany, such as Turkey and Poland. Remarkably, it is clear that such countries are not driving the results. In fact, with the naked eye, no clear pattern can be seen in the last two decades that suggests a strategic presence of the Goethe Institute in the direction of economically prospering regions that are particularly attractive for recruiting skilled workers. Although changes can be observed in China – which has become increasingly important in recent years in terms of emigration flows – there has been a significant shift in the Goethe Institute’s activity in various countries of the world, which have almost certainly played no role in Germany’s political considerations regarding the acquisition of skilled workers in the past, such as Mexico, Argentina and countries in the Middle East. These observations strengthen our confidence that the expansion agenda of the Goethe Institute does not systematically coincide with the migration policy of the German government.

Figure 2: Worldwide variation in the number of Goethe Institutes exploited for identification



Source: Authors' elaboration based on a dataset developed by Uebelmesser/Huber/Weingarten (2018). The map highlights countries exhibiting variation over time in the number of Goethe Institutes *with* language services. Such variation within countries is exploited in subsequent estimations.

Moreover, we empirically investigate whether there is a relationship between the opening or closing of language learning facilities abroad and potential economic and political variables that could, in turn, affect the self-selection of international migrants to Germany. Specifically, we regress the probability of opening or closing and the number of Goethe Institutes on a set of potential determinants (with a one-year lag) for the set of countries and the time period that corresponds to our subsequent main analyses. Because we apply panel fixed-effects estimation, the identification relies solely on within-country variation in variables. In addition to purely economic factors – countries' unemployment, GDP per capita and trade links with Germany – we include political variables regarding bilateral labor agreements, EU and Schengen membership and the Freedom House indicators of political rights and civil liberties (Freedom House, 2019) as well as for the migration history from the country to Germany (measured as the stock of migrants from the country in Germany). Finally, the number of casualties in war and armed conflict taken from UCDP Conflict Encyclopedia (2018) serves as an indicator of safety, which allows us to test whether safety considerations that may threaten, for example, employees of Goethe Institutes influence the location decisions.

Table 1 shows that endogeneity concerns regarding the utilization of language learning facilities as a strategic element of migration policy in the past are not supported by the data. None of the included factors are significant predictors of openings or closings of language institutes with two exceptions. Variations in GDP per capita in the origin country appear to be related to the opening and closing of Goethe Institutes, but not with the total number of institutes present in the country, which is the main variable used in our analysis.<sup>7</sup> An increase in the measure of political rights provided by the Freedom House organization, corresponding to a reduction in political rights, is significantly and negatively associated with the opening of new Goethe Institutes. Overall, this is consistent with the line of argument derived from the institutional background of Goethe Institutes, namely that their agenda is driven by political and cultural factors unrelated to the labor market or migration policy of the German government.

## 5.1 Descriptive statistics

Table 2 presents summary statistics of the main variables used in the analysis to describe the sample. Self-reported German skills at the time of immigration indicate that, on a scale from 1 to 5 where 1 is “no German at all” and 5 is “very good German skills”, immigrants in our sample tend to arrive with very poor skills on average. Only 16 percent of respondents enter with good or very good German skills, and 48 percent arrive with absolutely no knowledge of the German language. Just above 1 immigrant in 5 (21.8 percent) declares having taken German language classes in the country of origin.

With respect to gender, the sample used for analysis is slightly imbalanced in favor of women, who represent 54.1 percent of immigrants. On average, immigrants arrived as relatively young adults between 29 and 30 years old, although this variable displays con-

---

<sup>7</sup> Nevertheless, this underlines the importance of controlling for this important determinant of migration in all regressions.

Table 1: Fixed-Effects Estimation of Determinants of Location of GI

Estimator	OLS	OLS	OLS
	(1)	(2)	(3)
	Opening	Closing	Number of GI
Running sum of bilateral labor agreements	0.037 (0.330)	-0.037 (-0.830)	0.078 (0.556)
Unemployment rate	0.004 (0.409)	0.006 (0.660)	0.002 (0.135)
EU Country	-0.016 (-0.084)	-0.057 (-0.476)	0.048 (0.453)
Schengen Country	0.095 (0.829)	0.108 (0.883)	0.029 (0.359)
Log GDP p.c.	-0.396** (-2.088)	-0.408** (-1.997)	0.069 (0.412)
Share of migrants with spec. nationality in Germany	0.397 (0.092)	1.263 (0.329)	2.219 (0.395)
Share of German exports in year (EUR)	-0.010 (-0.101)	-0.106 (-0.762)	0.155 (1.375)
Number of casualties, hundreds	0.000 (0.557)	0.000* (1.679)	-0.000 (-0.698)
Freedom House: Political rights, 1 (High) - 7 (Low)	-0.071*** (-2.713)	-0.010 (-0.531)	-0.064** (-2.186)
Freedom House: Civil Liberties, 1 (High) - 7 (Low)	0.028 (0.731)	0.021 (0.732)	0.014 (0.397)
Constant	3.683** (2.191)	3.770** (2.049)	0.424 (0.262)
Observations	1,227	1,227	1,227
Countries	82	82	82
Years	2001-2014	2001-2014	2001-2014
Year FE	Yes	Yes	Yes
Country FE	Yes	Yes	Yes

Source: Authors' elaboration on data from Uebelmesser/Huber/Weingarten (2018), Freedom House (2019), UCDP Conflict Encyclopedia (2018), and Worldbank (2018). Notes: Dependent variables: Running sum over openings resp. closings of Goethe Institutes with language service (Models 1, 2). All independent variables enter with a one-year lag. Sample includes countries that are also part of the main regressions in Tables 3 and 4, except Syria and Cuba due to missing data. t statistics in parentheses. Standard errors clustered at country level. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

siderable variability because some migrants arrived at a much younger (5) or older (61) age.

The level of educational attainment is relatively low, with 49.4 percent of immigrants having no background in formal education. However, with 23.8 percent bringing a professional degree and one-quarter having attended university, approximately half of the sample still arrives with relevant human capital. They also bring valuable labor market experience, with 8.53 years of work experience on average prior to migration.

Economic conditions in both the origin countries and Germany exhibit considerable variability. Despite that, on average, the GDP per capita of migrants' country of origin was approximately 17,300 EUR in the year preceding arrival (PPP, 2011 USD), some migrants likely left their home country due to poor living conditions and severe recessions. With an average of 10.6 percent, the unemployment rate prior to migration was also relatively high. The nationality-specific monthly median wage was distributed between 426 and 3,076 euros, with an average of 1,389 euros per month.

Table 2: Summary statistics

	observations	mean	standard deviation	minimum	maximum
German skills (1-5 scale)	2,486	1.938	1.158	1	5
Good German skills	2,486	0.163	0.369	0	1
Some German skills	2,486	0.519	0.500	0	1
Took German classes	2,486	0.218	0.413	0	1
Number of language institutes in origin country, immigration year(t-1)	2,486	1.577	1.044	0	7
Gender (1= Female)	2,486	0.541	0.498	0	1
Age at immigration	2,486	29.661	10.681	5	61
No education	2,486	0.494	0.500	0	1
On-the-job training	2,486	0.018	0.133	0	1
Professional degree	2,486	0.238	0.426	0	1
University degree	2,486	0.235	0.424	0	1
Doctoral studies	2,486	0.014	0.119	0	1
Work experience before migration, in years	2,486	8.530	9.354	0	44
Job offer at arrival	2,486	0.105	0.306	0	1
GDP per capita in origin country, immigration year(t-1)	2,486	17,372	9,747	618	64,586
Unemployment rate in origin country, immigration year(t-1)	2,486	10.644	5.176	1	37
Median wage in Germany by nationality, immigration year	2,486	1,389	421.5	426	3,076

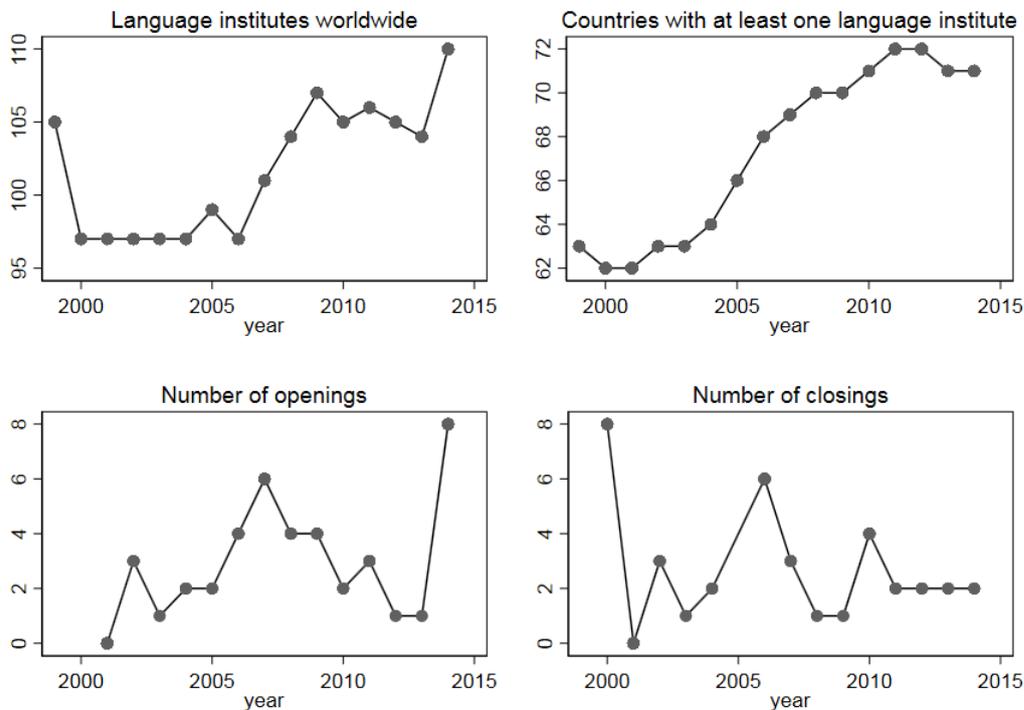
Source: Authors' elaboration on IAB-SOEP Migration Sample and Uebelmesser/Huber/Weingarten (2018).

There were 149 institutes operating in 94 countries in 2014 (Uebelmesser/Huber/Weingarten, 2018). In the sample used in our analysis, which includes 84 countries, the number of observed institutes with language services per origin country and year varies between 0 and 7, with an average of 1.58 institutes. Although the number of Goethe Institutes increased over our period of analysis, Figure 3 highlights two important features of the variable. First, the number of operating institutes varies substantially between years because some institutes are closed while others are opened in various countries or cities in a given country.<sup>8</sup>

<sup>8</sup> We do not rely on spatial variation in the presence of language learning facilities within countries because we do not have sufficient information on the locations of immigrants within their countries of origin prior to migration.

Second, the number of countries with at least one institute offering language services increased but not steadily. This entails that the exposure to institutes with language services for different cohorts varies along both the intensive and extensive margins.

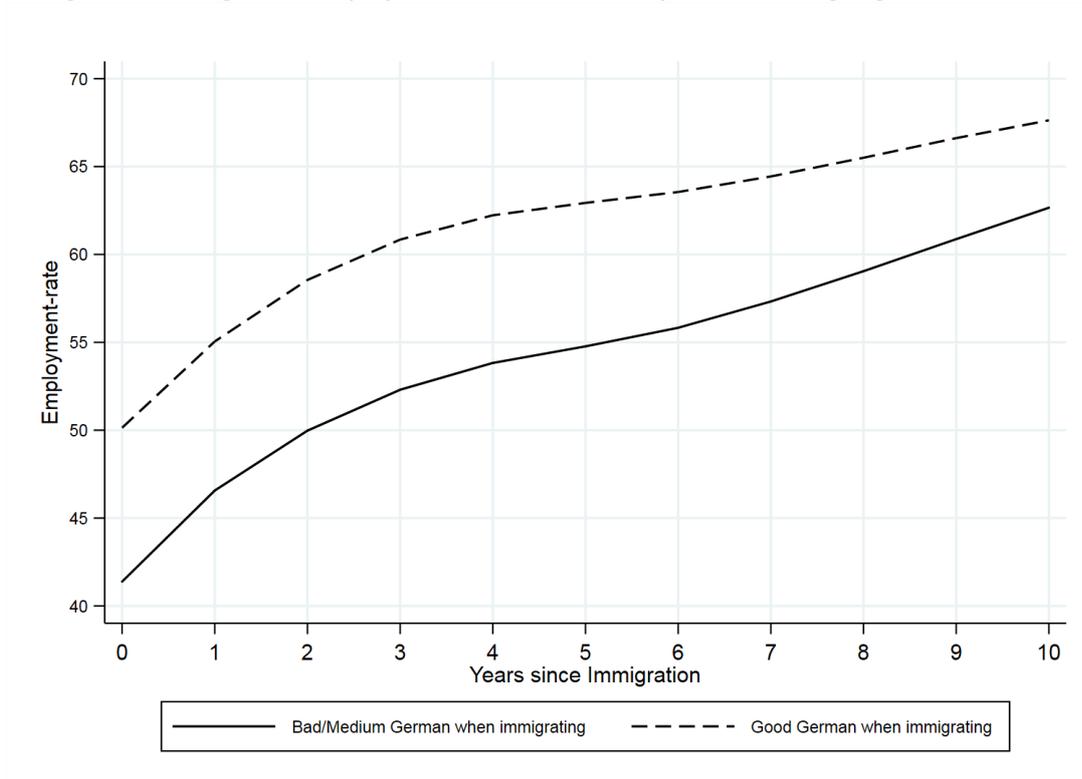
Figure 3: Number of Goethe Institutes with language services over time



Source: Authors' elaboration based on a dataset developed by Uebelmesser/Huber/Weingarten (2018). The y-axis in each sub-figure represents the number of Goethe Institutes *with* language services described by the title of the sub-graph. Goethe Institutes without language services are not considered in the figure.

The aim of our analysis is to understand whether exposure to German language learning institutes enhances immigrants' language skills at arrival and, if so, whether the underlying mechanisms involve shifts in selection patterns along other characteristics. Destination country language skills are among the most important determinants of the integration potential of newly arrived migrants. Indeed, beyond a direct effect of literacy on productivity, language may also facilitate the transfer of origin-specific human capital to the destination country labor market. Figure 4 offers some suggestive evidence supporting this argument. Based on retrospective calendar data, it plots employment rates as a function of years since immigration to Germany for migrants who arrived with good or very good German skills against migrants who arrived with poor German language skills or no German skills at all. Although the interpretation cannot be causal, the results suggest a strong relationship between language skills at immigration and initial labor market performance. Furthermore, despite a catch-up effect of the group with no language skills at arrival, the initial advantage appears to be sustained for a relatively long time period.

Figure 4: Immigrants' employment rate over time by German language skills at arrival



Source: Authors' elaboration on GSOEP panel data, smoothed. First-time interviewees. The sample is the same as that in the multivariate regressions in Table 3. Notes: Full- or part-time employment. Good language skills are defined as a mean score for speaking, writing, and reading equal to four or higher on a 1-5 scale. Mean scores below four are classified as poor language skills.

## 5.2 Main results

The main results of our analysis are reported in Table 3. Column (1) reports a statistically significant effect of the number of German language learning facilities in the year before immigration on language skills at arrival. Specifically, holding all other variables fixed, one additional German language learning facility in the country of origin leads to an average improvement in German language skills of 0.125 points on a 1 to 5 scale.<sup>9</sup>

The results also suggest that women in the sample we use for analysis bring better German language skills on average than men. In contrast, holding all other variables fixed, age at immigration does not appear to make a significant difference with respect to German language skills at arrival. Interestingly, neither economic conditions in the country of origin nor nationality-specific median wages in Germany significantly affect German language skills at arrival. This suggests that, while macroeconomic conditions might shape migration flows by modifying the returns to migration, the selection patterns induced by such modified incentives to migrate do not affect the acquisition of destination country language skills. The latter are instead determined by individual-level characteristics such as level of education and exposure to language learning opportunities.

Columns (2) and (3) report the effect of the number of German language learning facilities on the probability of having German language skills at immigration above a specific threshold. The dependent variable in column (2) is equal to 1 if the immigrants had “good” or “very good” German skills at arrival and 0 otherwise. In column (3), the dependent variable takes value 1 if the immigrants had “very good”, “good” or just “some” German skills at arrival and 0 otherwise. The results indicate that the increased presence of learning opportunities improves language skills at different points of the distribution. Moreover, the effect is almost equal at lower and upper levels of language skills.

Next, we investigate whether the effect of language institutes varies for different types of language skills, such as speaking, writing, and reading. Indeed, it is plausible that reading and writing might be easier to learn abroad than speaking, which would require more intensive communication with natives on a regular basis. The results in columns (4), (5) and (6) suggest that the effect of language institutes affects all types of language skills, although the effect on speaking skills does appear to be slightly lower than the effect on writing and reading skills.

## 5.3 Effects on other characteristics

The findings above raise the question of the underlying channel through which language learning institutes improve language skills at immigration. The most obvious channel is through participation in language classes in the country of origin. However, the effect could also operate through other channels because the opening of a language learning institute

---

<sup>9</sup> The results are robust to controlling for the political stability in the source country by including the Freedom House indicators of political rights and civil liberties (Freedom House, 2019), which could correlate with both variation in the presence of language services and self-selection of migrants. These additional results are available upon request from the authors.

modifies many characteristics in the environment of potential migrants. As discussed in Section 3, a standard self-selection model suggests that the presence of language learning services abroad raises migration incentives disproportionately more for persons with high levels of education. This effect could, in principle, operate through several channels beyond participation in language classes. For instance, Goethe Institutes might provide highly motivated and skilled potential migrants with books and media in German, allowing the autonomous acquisition of language skills. Employees of such institutes might also form networks with the local population, fostering informal learning. Goethe Institutes might also modify the information structure available to potential migrants by providing crucial information about the labor market and living conditions in Germany. In turn, such additional information might induce potential migrants to update their preferences. As a result, some individuals who would have otherwise chosen a different country might consider Germany as a potential destination. To the extent that such candidates are drawn from a different part of the education distribution in the country of origin, this could also indirectly influence the language skill composition of a given cohort.

Table 4 presents the results. Column (1) confirms the most intuitive channel by showing that language learning institutes do significantly influence the probability of taking German classes prior to migration. However, as shown in column (2), the number of language learning institutes in the country of origin also significantly enhances the likelihood that an immigrant has higher education, which is defined as having attended at least some university. Similarly, column (3) reports a positive and statistically significant effect on the number of years of labor market experience prior to migration. These two effects could be attributable to the channel through which people can immigrate in Germany, which determines their status at immigration. Indeed, the role of language requirements varies with the type of visa requested for immigration. For example, a visa for labor immigration requires a good command of German language, while visas for family reunification or asylum do not have language skill requirements. Hence, the opportunity to learn German in the home country would give potential migrants access to this additional channel at a reduced cost.

The results in column (4) indeed show that the number of language learning institutes in the country of origin is associated with a significantly higher probability that an immigrant will arrive through the labor migration channel. Taken together, the results suggest that the availability of language learning opportunities abroad affects several characteristics beyond improved language skills.

## **5.4 Robustness checks and sensitivity analyses**

In this section, we present additional tables challenging the validity and interpretation of our main results with respect to several threats to identification.

### **5.4.1 Goethe Institutes without language services**

First, we address the possibility of spurious correlation between the opening of Goethe Institutes and the outcome variables. We exploit the fact that not all Goethe Institutes offer

Table 3: Effect of language institutes on language skills at immigration

Estimator	OLS (1)	Probit (2)	Probit (3)	OLS (4)	OLS (5)	OLS (6)
VARIABLES	German skills (1-5 scale)	Good German skills	Some German skills	Speaking skills	Writing skills	Reading skills
Number of language institutes in origin country, immigration year(t-1)	0.125*** (0.034)	0.049*** (0.015)	0.045** (0.018)	0.117*** (0.036)	0.136*** (0.033)	0.122*** (0.039)
Gender = Female (reference is male)	0.127** (0.050)	0.047*** (0.016)	0.011 (0.017)	0.077 (0.048)	0.146*** (0.052)	0.157*** (0.055)
Age at immigration	0.004 (0.004)	-0.000 (0.001)	0.003* (0.001)	0.005 (0.004)	0.003 (0.004)	0.003 (0.004)
Income before immigration	-0.000*** (0.000)	-0.000 (0.000)	-0.000** (0.000)	-0.000*** (0.000)	-0.000*** (0.000)	-0.000*** (0.000)
Log GDP per capita in origin country, immigration year(t-1)	-0.039 (0.276)	-0.189* (0.105)	0.013 (0.147)	-0.040 (0.276)	-0.108 (0.281)	0.030 (0.296)
Unemployment rate in origin country, immigration year(t-1)	-0.013 (0.013)	-0.004 (0.003)	-0.008* (0.005)	-0.010 (0.013)	-0.014 (0.013)	-0.014 (0.013)
Median wage in Germany by nationality, immigration year	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.001* (0.000)
Origin country joined EU	-0.253 (0.190)	-0.040 (0.052)	-0.135* (0.069)	-0.208 (0.190)	-0.280 (0.179)	-0.272 (0.205)
Observations	2,486	2,486	2,486	2,486	2,486	2,486
R-squared	0.155			0.148	0.146	0.157
Origin fixed effects	YES	YES	YES	YES	YES	YES
Immigration year fixed effects	YES	YES	YES	YES	YES	YES

Source: Authors' elaboration on IAB-SOEP Migration Sample and Uebelmesser/Huber/Weingarten (2018). Notes: All specifications also control for relationship status before migration. Standard errors (in parentheses) are clustered at the origin country level. Marginal effects are reported for probit models (columns 2 and 3). Statistical significance levels: \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

Table 4: Transmission channels of the effect of language institutes

Estimator	Probit (1)	Probit (2)	OLS (3)	Probit (4)
VARIABLES	Took German classes	Some higher education	Pre-migration labor market experience	Job offer at arrival
Number of language institutes in origin country, immigration year(t-1)	0.042** (0.018)	0.033*** (0.012)	0.679*** (0.120)	0.032** (0.014)
Gender = 2, Female (reference is male)	0.024 (0.017)	0.018 (0.017)	-2.394*** (0.339)	-0.106*** (0.022)
Age at immigration	0.000 (0.001)	0.015*** (0.001)	0.736*** (0.028)	0.001 (0.001)
Work experience before immigration, in years	-0.003** (0.001)	-0.016*** (0.002)		0.001 (0.001)
Income before immigration	-0.000*** (0.000)	-0.000 (0.000)	0.000*** (0.000)	-0.000*** (0.000)
Log GDP per capita in origin country, immigration year(t-1)	-0.008 (0.133)	-0.128 (0.102)	-1.125 (1.317)	0.014 (0.124)
Unemployment rate in origin country, immigration year(t-1)	0.001 (0.005)	-0.002 (0.004)	-0.009 (0.036)	-0.001 (0.003)
Median wage in Germany by nationality, immigration year	-0.000* (0.000)	0.000 (0.000)	0.000 (0.001)	0.000 (0.000)
Education abroad (ref. "No formal education")				
On-the-job training	0.038 (0.067)		0.087 (0.811)	0.043 (0.051)
Professional degree	0.022 (0.026)		-0.369 (0.454)	0.037 (0.033)
University studies	0.166*** (0.028)		-2.703*** (0.319)	0.048** (0.021)
Doctoral studies	0.250*** (0.090)		-4.943*** (0.572)	0.112* (0.065)
Origin country joined EU	-0.079 (0.076)	-0.101*** (0.035)	0.312 (0.361)	0.009 (0.040)
Observations	2,486	2,486	2,486	2,486
R-squared			0.734	
Origin fixed effects	YES	YES	YES	YES
Immigration year fixed effects	YES	YES	YES	YES

Source: Authors' elaboration on IAB-SOEP Migration Sample and Uebelmesser/Huber/Weingarten (2018). Notes: All specifications also control for relationship status before migration. Standard errors (in parentheses) are clustered at the origin country level. Marginal effects are reported for probit models (columns 1, 2, and 4). Statistical significance levels: \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

language courses. While this is only the case for a minor share, approximately 15 % in 2014, we can still use these facilities in a placebo test to challenge our previous findings. If the effect of Goethe Institutes on language skills at immigration is truly attributable to the availability of language services, we would expect no significant effects of changes in the number of institutes without language services, at least on the variables *German skills* and *course participation*. Columns (1) and (3) of Table 5 show that the opening of a Goethe Institute without language services is not sufficient to improve German skills at arrival, suggesting that the provision of language services is key. Nevertheless, positive spillover effects such as network effects or access to information on non-language variables might still be expected from institutes that do not offer language courses. However, we do not find increased shares of higher education among immigrants with higher exposure (column 5).

#### 5.4.2 Demand for German language classes

Additionally, we perform falsification tests using the one-period-ahead number of Goethe Institutes as the main explanatory variable. This test addresses the legitimate concern of reverse causality: Goethe Institutes might open in countries to meet an increasing demand for German language classes, which could be driven by rising incentives to migrate to Germany.<sup>10</sup> In this case, we would expect the variation in outcome variables such as German skills at arrival, the probability of taking a German class, or education level at arrival to precede the opening of Goethe Institutes. Hence, the outcome variable in immigration year  $t$  would be strongly correlated with the variable of interest in year  $t + 1$ . On the contrary, if we truly measure the effect of differences in exposure to language services of individuals in our main regressions in Tables 3 and 4, then the number of institutes one year *after* immigration should not be significantly associated with the outcome variables. This interpretation is supported by the results in columns (2), (4) and (6) of Table 5, suggesting that our main results are not demand driven due to increased migration incentives to Germany.

#### 5.4.3 Heterogeneity of main results

To gain more precise insights into the effect of language learning facilities, we conduct sensitivity analyses in Table 6 with respect to (i) the linguistic proximity between the language of immigrants' countries of origin and German, (ii) gender and (iii) GDP per capita in immigrants' countries of origin. In Section 3, we argued that language learning facilities abroad reduce the opportunity costs of language acquisition. Correspondingly, persons from countries with a high linguistic distance can be expected to particularly profit from such services, as they face greater learning effort and higher costs of language acquisition

---

<sup>10</sup> Recall that our specification includes the median wage in Germany by nationality, log GDP per capita in the origin country in the year prior to immigration, and the unemployment rate in the origin country in the year prior to immigration. These control variables already capture variations in origin-specific incentives to migrate to Germany. Nevertheless, it is important to further investigate this threat to identification through alternative specifications.

Table 5: Effect of Goethe Institutes *without* language services and one year after immigration

Estimator	OLS (1)	OLS (2)	Probit (3)	Probit (4)	Probit (5)	Probit (6)
VARIABLES	German skills 1-5 scale	German skills 1-5 scale	Took German classes	Took German classes	Some higher education	Some higher education
GI without language service, immigration year(t-1)	0.059 (0.083)		0.006 (0.021)		0.009 (0.011)	
GI with language service, immigration year(t+1)		0.042 (0.049)		0.021 (0.029)		0.021 (0.019)
Observations	2,486	2,456	2,486	2,456	2,486	2,456
R-squared	0.154	0.154				
Origin fixed effects	YES	YES	YES	YES	YES	YES
Immigration year fixed effects	YES	YES	YES	YES	YES	YES

Source: Authors' elaboration on IAB-SOEP Migration Sample and Uebelmesser/Huber/Weingarten (2018). Notes: Regressions include control variables for participation in German classes abroad, number of language institutes in the origin country in year prior to immigration, gender, age at immigration, relationship status before migration, work experience before immigration, education level before migration, income before migration, median wage in Germany by nationality, log GDP per capita in the origin country in year prior to immigration, and the unemployment rate in the origin country in the year prior to immigration. Standard errors (in parentheses) are clustered at the origin country level. Marginal effects are reported for probit models (columns 3, 4, 5, and 6). Statistical significance levels: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

(Isphording/Otten, 2014). Therefore, in column (1), we test for heterogeneity in the effect of Goethe Institutes on German language skills at arrival.<sup>11</sup>

Contrary to expectations, the non-significant interaction in column (1) provides evidence for a linear effect of language learning facilities along the distribution of origin countries' linguistic proximity. One explanation for this result is that average language skills at the time of moving to Germany are poor in any case (Section 5.1). We can thus assume that in most cases, virtually no prior knowledge of German is available and that Goethe Institutes abroad teach those planning to migrate to Germany very basic skills. If linguistic proximity reduces the required language learning effort more at higher levels of language fluency, when more complex grammatical structures are learned, such initial, basic learning effects may be more or less independent of the linguistic distance of the mother tongue. Second, in view of the composition of the origin countries shown on the world map in Figure 2, our sample contains little variation in linguistic proximity because linguistically distant countries are over-represented; the median linguistic proximity of origin-countries in our estimates is approximately 0.125 on the scale from 0 (furthest away) to 0.881 (closest to German), leaving little variation to identify the effect of linguistic distance.

Column (2) reveals a significant improvement in German language skills at the time of immigration for females relative to males. Studies suggest that women might be more efficient in acquiring a second language (Van Der Slik/Van Hout/Schepens, 2015). In addition, although not verifiable with our data, it can be assumed that the possibility of language acquisition in the country of origin prior to emigration reduces opportunity costs more for women – for example, through higher care opportunities for children within the family during course participation. Finally, column (3) shows a stronger effect of language institutes on language skills for economically poorer countries as measured by GDP per capita. This could be because in less prosperous countries – e.g., due to a low provision of public education – there are fewer alternative opportunities to acquire the German language, so the

<sup>11</sup> Note that the regressions do not include the non-interacted linguistic proximity due to multicollinearity with the origin country fixed effects.

Goethe Institute is particularly relevant in such contexts. This result is also consistent with opportunity costs as a mechanism for the effect: as living expenses and forgone earnings are lower in poor countries, the gains from learning German abroad (compared to learning German upon arrival) are higher in poor countries.

Table 6: Sensitivity analyses of Goethe Institutes

Estimator	OLS (1)	OLS (3)	OLS (4)
VARIABLES	German skills 1-5 scale	German skills 1-5 scale	German skills 1-5 scale
Number of language institutes in origin country, immigration year(t-1)	0.249* (0.131)	0.093*** (0.033)	0.106*** (0.023)
Female	0.096** (0.048)	-0.011 (0.084)	0.096** (0.048)
GDP p.c. origin country (t-1)			
1st quartile (ref. 4th quartile)	-0.489 (0.395)	-0.479 (0.394)	-0.824* (0.482)
2nd quartile	-0.235 (0.277)	-0.227 (0.280)	-0.485** (0.238)
3rd quartile	-0.005 (0.104)	-0.005 (0.103)	0.042 (0.193)
<i>Interactions with number of language institutes (t-1)</i>			
linguistic proximity	-0.687 (0.729)		
Female		0.065* (0.037)	
GDP 1st quartile (ref. 4th quartile)			0.472** (0.237)
GDP 2nd quartile			0.441** (0.221)
GDP 3rd quartile			-0.028 (0.104)
Observations	2,478	2,486	2,486
R-squared	0.167	0.169	0.169
Origin fixed effects	YES	YES	YES
Immigration year fixed effects	YES	YES	YES

Source: Authors' elaboration on IAB-SOEP Migration Sample and Uebelmesser/Huber/Weingarten (2018). Regressions include control variables for age at immigration, relationship status before migration, work experience before immigration, education level before migration (except column 1), income before migration, median wage in Germany by nationality, and the unemployment rate in the origin country in the year prior to immigration. Standard errors (in parentheses) are clustered at the origin country level. Statistical significance levels: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

## 6 Mediation analysis

Our empirical results indicate that immigrants' German language skills at arrival are strongly influenced by the presence of German language learning facilities prior to immigration. However, our results also show that exposure to German language learning facilities in the country of origin induces self-selection in other aspects not directly related to participation in language classes. This suggests that the effect of German language learning facilities on language skills at arrival could be driven by other factors beyond participation in formal classes. For instance, the presence of Goethe Institutes could provide access to books

Figure 5: Illustration of a mediated effect



and media content in German, allowing highly motivated potential migrants to acquire basic German skills by themselves. Furthermore, the presence of German nationals working in the institutes could also induce informal learning effects beyond the formal classes. If a considerable share of the effect of German language learning facilities on language skills at arrival is mediated through channels other than participation in language classes, this would raise important questions regarding the factors driving the choice to learn a foreign language and the means used to achieve the language learning objective. Furthermore, strong indirect effects would provide evidence on the effectiveness of language learning facilities as a facilitator of the international migration process by making migration to Germany accessible to a pool of highly motivated and qualified potential migrants. Therefore, a better understanding of the transmission channels of the presence of German language learning facilities is highly relevant from both an academic and a policy perspective.

To further explore whether the presence of German language learning facilities affects immigrants' German language skills at arrival through channels other than formal language classes, we propose to explore the mediation effects.

## 6.1 Mechanism

We seek to reveal the mechanism driving the effects of the presence of German language learning facilities prior to immigration on immigrants' German language skills at arrival. Specifically, we focus on participation in German language classes in the country of origin as a mediating variable. As illustrated in Figure 5, formal mediation analysis allows us to decompose the total effect of our treatment variable  $GI$ , the number of language learning facilities in the country of origin, on the outcome  $Y$ , German language skills at arrival, into two effects: the average causal mediation effect (ACME), which operates through the mediator variable  $M$ , participation in German language classes in the country of origin, and the average direct effect (ADE), which combines the remaining effects (i.e., the unmediated effects). Note that the ADE includes effects transmitted through all other channels such as self-selection in ability and informal learning.

Formally, we consider the following set of linear equations:

$$M_{iot} = \alpha_2 + \beta_4 GI_{o(t-1)} + \beta_5 X_i + \beta_6 Z_{o(t-1)} + \delta_o + \delta_t + \varepsilon_{iot} \quad (2)$$

$$Y_{iot} = \alpha_3 + \beta_7 GI_{o(t-1)} + \beta_8 M_{iot} + \beta_9 X_i + \beta_{10} Z_{o(t-1)} + \delta_o + \delta_t + \varepsilon_{iot} \quad (3)$$

where  $Y_{iot}$  is an individual-level outcome variable measuring the characteristics of immigrant  $i$  originating from country  $o$  who arrived in Germany in year  $t$ .  $M_{iot}$  is a mediator,

referring to participation in German language classes in the country of origin.  $GI_{o(t-1)}$  represents the number of language facilities present in country of origin  $o$  one year before migration, as our treatment variable. We also introduce  $X_i$ , a set of time-invariant socioeconomic variables measured at the time of immigration and including gender, age at immigration, work experience prior to immigration (in years), and the level of education attained in the country of origin.  $Z_{o(t-1)}$  is a set of macroeconomic variables capturing the economic conditions in both the country of origin one year prior to migration and Germany. We use GDP per capita and the unemployment rate as proxies for economic conditions in the country of origin. For each country of origin, median wages in Germany are computed for workers holding the corresponding nationality for each immigration year  $t$  to capture economic conditions at immigration. The sets of fixed effects  $\delta_o$  and  $\delta_t$  control for time-invariant origin country characteristics and origin-invariant immigration year characteristics, respectively. The terms  $\alpha_2$  and  $\alpha_3$  refer to the intercepts, while  $\varepsilon_{2iot}$  and  $\varepsilon_{3iot}$  indicate the idiosyncratic error terms.

The mediation analysis approach proposed by Imai/Keele/Yamamoto (2010) consists of a four-step algorithm. In the first step, one model is fit for both the observed outcome and mediator variables. Second, the model parameters are simulated from their sampling distribution.<sup>12</sup> Third, all potential values of the mediator are modeled conditional on treatment and covariates. Finally, summary statistics, point estimates, and confidence intervals are computed.

This methodology produces an unbiased estimate of the ACME under very clear assumptions. Hence, the advantage of using the framework designed by Imai/Keele/Yamamoto (2010) is the transparency regarding the underlying identifying assumption of sequential ignorability and the ability to conduct formal sensitivity analysis. Sequential ignorability relies on several strong assumptions with severe implications.

The first assumption states that treatment status is ignorable, conditional on covariates. In other words, there are no unobserved confounding variables that affect both the number of language facilities present in the country of origin ( $GI_{o(t-1)}$ ) and our outcome of interest ( $Y_{iot}$ ) or the mediator ( $M_{iot}$ ). In our case, the country of origin fixed effects  $\delta_o$  combined with the institutional setting ensure the plausibility of such a condition. The second assumption entails that the mediator is ignorable, conditional on treatment status and covariates. Specifically, this condition requires that no unobserved variables affect both the considered outcome ( $Y_{iot}$ ) and the mediator ( $M_{iot}$ ) after conditioning on observable variables that do affect them. We argue that the setting of our analysis makes this assumption plausible: conditional on gender, age at immigration, education level, and labor market experience, no unobserved variable should affect both German language skills at arrival and participation in German language classes in the country of origin. Although it cannot be formally tested with the observed data, the robustness of our empirical results to potential violation of this assumption will be evaluated using the procedure developed by Imai/Keele/Yamamoto (2010).

<sup>12</sup> We use 1000 simulations. See Imai/Keele/Yamamoto (2010) for technical details on the algorithm.

## 6.2 Transmission channels

Table 7 reports the results of the mediation analysis. Column (1) refers to Equation 2 and shows the impact of the number of language facilities present in the country of origin on the mediator. Column (2) refers to Equation 3 and displays the impact of the number of language facilities present in the country of origin and participation in German language classes in the country of origin on German language skills at arrival. The lower part of the table presents the ACME with 95 percent confidence intervals, along with the direct effect, the total effect, and the percentage of the total effect mediated through participation in language classes in the country of origin. In line with expectations based on results from Section 5.2, we find that the presence of additional language facilities in the country of origin one year prior to migration significantly increases the probability of participating in German language classes before migration. Consistently, we find that participation in German language classes in the country of origin is positively and significantly associated with better German language skills at arrival. These results suggest that the effect of Goethe Institutes on German language skills at immigration is indeed mediated through participation in German language classes. However, previous results from Section 5.2 suggest spillover effects from the number of language learning facilities on the self-selection of immigrants along other characteristics such as education level that could also be associated with German language skills at arrival. The results presented thus far are not informative with respect to the relative importance of different channels.

Regarding participation in German language classes as a transmission channel, we have an ACME  $\approx 0.034$ .<sup>13</sup> The mediated effect as a percentage of the total effect of the number of language facilities present in the country of origin corresponds to the ratio of ACME over the predicted coefficient for the number of language learning facilities present in the country of origin and is equal to 0.132. Accordingly, approximately 25.5 percent ( $\approx \frac{0.034}{0.132}$ ) of the total effect of language learning facilities operated by the Goethe Institute present in the country of origin is mediated through participation in German classes. The remaining direct effect, representing approximately 3/4 of the total effect, might be explained by different self-selection patterns induced by the presence of language learning facilities or by the informal learning opportunities they generate.

The estimates in Table 7 are only valid to the extent that the sequential ignorability assumptions hold. While the assumptions cannot be tested, it is possible to assess the sensitivity of the results to violations of the underlying assumptions. To do so, the correlation  $\rho$  between the error terms of Equations 2 and 3 is computed. Note that the sequential ignorability assumptions imply that  $\rho = 0$ . Hence, the sensitivity test reduces to assessing the maximum value of  $\rho$  for which the results would still be valid.<sup>14</sup> Figure 6 plots  $\rho$  against the estimated ACME. The results indicate that the conclusions regarding the direction of the ACME would be valid as long as  $\rho$  does not exceed 0.358, implying that the results of the mediation analysis are relatively robust to violations of the underlying assumptions.

<sup>13</sup> Note that ACME is the product of the effect of the number of language facilities present in the country of origin on the considered mediator and the effect of the mediator on German language skills at arrival. Based on the figures in Table 7, the 3-digit ACME is  $0.034 \approx 0.037 \times 0.977$ .

<sup>14</sup> For technical details, see Imai/Keele/Yamamoto (2010).

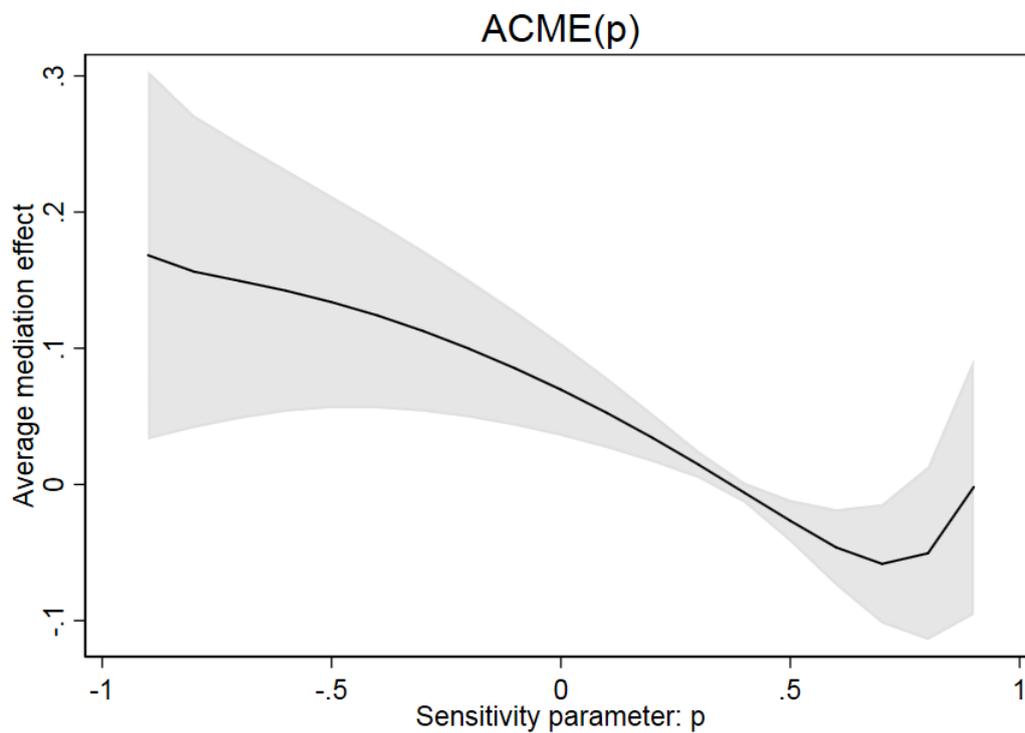
Table 7: Mediation analysis: direct and indirect effects of language institutes abroad

Estimator	OLS (1)	OLS (2)
VARIABLES	Took German classes	German skills (1-5 scale)
Took German classes abroad		0.977*** (0.072)
Number of language institutes in origin country, immigration year(t-1)	0.037*** (0.014)	0.091** (0.039)
Gender	0.018 (0.017)	0.076 (0.046)
Age at immigration	-0.000 (0.001)	0.008 (0.006)
Work experience before immigration, in years	-0.003** (0.001)	-0.004 (0.006)
Observations	2,486	2,486
R-squared	0.121	0.274
Education abroad fixed effects	YES	YES
Origin fixed effects	YES	YES
Immigration year fixed effects	YES	YES

Mediation analysis		
Effect	Mean	[95% Conf. Interval]
ACME	0.034	-0.012 0.080
Direct Effect	0.098	-0.017 0.220
Total Effect	0.132	0.006 0.261
Percentage of Total Effect mediated	0.255	0.118 2.086

Source: Authors' elaboration on IAB-SOEP Migration Sample and Uebelmesser/Huber/Weingarten (2018). Notes: Regressions also include control variables for relationship status before migration, education level before migration, income before migration, median wage in Germany by nationality, log GDP per capita in the origin country in the year prior to immigration, and the unemployment rate in the origin country in the year prior to immigration. The ACME is the average causal mediation effect. Standard errors in parentheses \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

Figure 6: Sensitivity of estimated indirect effects of language institutes abroad



Source: Authors' elaboration on IAB-SOEP Migration Sample and Uebelmesser/Huber/Weingarten (2018). Notes: regressions include control variables for participation in German classes abroad, the number of language institutes in the origin country in the year prior to immigration, gender, age at immigration, relationship status before migration, work experience before immigration, education level before migration, income before migration, median wage in Germany by nationality, GDP per capita growth rate in the origin country in the year prior to immigration, and the unemployment rate in the origin country in the year prior to immigration.

## 7 Conclusions

Language serves as a central determinant of migrants' successful integration into host societies and labor markets. Additionally, from the perspective of receiving countries, the language proficiency and qualification levels of migrants are highly relevant in view of demographic change and skills shortages. The aim of this paper was to estimate the effect of the presence of German language learning facilities in the country of origin on immigrants' language skills at arrival. In the analysis, transmission channels received particular attention. Indeed, we found that the presence of Goethe Institutes not only facilitates access to German learning opportunities but that it also modifies the information and network structure in the environment of potential migrants. We rely on a self-selection model with heterogeneous migration costs to discuss the consequences in terms of self-selection of migrants.

The empirical analysis is implemented on a sample of 2,486 immigrants originating from 84 countries who arrived in Germany between 2000 and 2014. Individual-level information taken from the IAB-SOEP Migration Sample is merged with data on the number of German language learning facilities in immigrants' country of origin obtained from Uebelmesser/Huber/Weingarten (2018).

Our results show that exposure to German language learning facilities improves self-reported language skills at arrival. However, we also find that the number of Goethe Institutes present in the country of origin prior to migration improves selection in terms of education and labor market experience. Consistently, immigrant cohorts with higher exposure to language learning institutes are more likely to immigrate with a work visa. A formal mediation analysis reveals that 25.5 percent of the total effect of exposure to German language learning institutes on language skills at immigration is mediated through actual participation in language classes abroad. This suggests that such language learning facilities bear important spillover effects and contribute to overcoming migration barriers through several channels.

In the context of the new German immigration law, the results of our analysis suggest that the strategy of the German government to increase its support for language learning institutes abroad is indeed an effective way to reconcile high language requirements for labor immigration with providing migration incentives for high-skilled migrants abroad.

## References

- Bauer, Thomas; Pereira, Pedro T.; Vogler, Michael; Zimmermann, Klaus F. (2002): Portuguese Migrants in the German Labor Market: Selection and Performance. In: *The International Migration Review*, Vol. 36, No. 2, p. 467–491, URL <http://www.jstor.org/stable/4149461>.
- Berman, Eli; Lang, Kevin; Siniver, Erez (2003): Language-skill complementarity: returns to immigrant language acquisition. In: *Labour Economics*, Vol. 10, No. 3, p. 265 – 290, URL <http://www.sciencedirect.com/science/article/pii/S0927537103000150>.
- Bertoli, Simone (2010): Networks, Sorting and Self-Selection of Ecuadorian Migrants. In: *Annals of Economics and Statistics*, , No. 97/98, p. 261–288, URL <http://www.jstor.org/stable/41219118>.
- Bleakley, Hoyt; Chin, Aimee (2004): Language Skills and Earnings: Evidence from Childhood Immigrants\*. In: *The Review of Economics and Statistics*, Vol. 86, No. 2, p. 481–496, URL <https://doi.org/10.1162/003465304323031067>.
- Borjas, George J. (2014): *Immigration Economics*. Cambridge, MA: Harvard University Press.
- Borjas, George J. (2008): Labor Outflows and Labor Inflows in Puerto Rico. In: *Journal of Human Capital*, Vol. 2, No. 1, p. 32–68, URL <https://doi.org/10.1086/527521>.
- Borjas, George J. (1994): The Economics of Immigration. In: *Journal of Economic Literature*, Vol. 32, No. 4, p. 1667–1717, URL <http://www.jstor.org/stable/2728791>.
- Borjas, George J. (1991): *Immigration and Self-Selection*. University of Chicago Press, p. 29–76, URL <http://www.nber.org/chapters/c6663>.
- Borjas, George J. (1987): Self-Selection and the Earnings of Immigrants. In: *The American Economic Review*, Vol. 77, No. 4, p. 531–553, URL <http://www.jstor.org/stable/1814529>.
- Brücker, Herbert; Bertoli, Simone; Facchini, Giovanni; Mayda, Anna Maria; Peri, Giovanni (2012): *Brain Drain and Brain Gain: The Global Competition to Attract High-Skilled Migrants: Part I Understanding Highly Skilled Migration in Developed Countries*. Oxford University Press.
- Brücker, Herbert; Kroh, Martin; Bartsch, Simone; Goebel, Jan; Kühne, Simon; Liebau, Elisabeth; Trübswetter, Parvati; Tucci, Ingrid; Schupp, Jürgen (2014): The new IAB-SOEP Migration Sample: an introduction into the methodology and the contents. In: *SOEP Survey Papers, Series C - Data Documentations*, Vol. 216.
- Budria, Santiago Rodriguez; Swedberg, Pablo (2015): The Impact of Language Proficiency on Immigrants ' Earnings. In: *Revista de Economía Aplicada*, Vol. 23, No. 67, p. 62–91.
- Chiquiar, Daniel; Hanson, Gordon H. (2005): International Migration, Self-Selection, and the Distribution of Wages: Evidence from Mexico and the United States. In: *Journal of Political Economy*, Vol. 113, No. 2, p. 239–281.

Chiswick, Barry R. (1998): Hebrew language usage: Determinants and effects on earnings among immigrants in Israel. In: *Journal of Population Economics*, Vol. 11, No. 2, p. 253–271, URL <https://doi.org/10.1007/s001480050068>.

Chiswick, Barry R. (1991): Speaking, Reading, and Earnings among Low-Skilled Immigrants. In: *Journal of Labor Economics*, Vol. 9, No. 2, p. 149–170, URL <http://www.jstor.org/stable/2535238>.

Chiswick, Barry R. (1978): The Effect of Americanization on the Earnings of Foreign-born Men. In: *Journal of Political Economy*, Vol. 86, No. 5, p. 897–921.

Chiswick, Barry R.; Miller, Paul W. (2003): The complementarity of language and other human capital: immigrant earnings in Canada. In: *Economics of Education Review*, Vol. 22, No. 5, p. 469 – 480, URL <http://www.sciencedirect.com/science/article/pii/S0272775703000372>, special Issue In Honor of George Psacharopoulos.

Chiswick, Barry R.; Miller, Paul W. (1995): The Endogeneity between Language and Earnings: International Analyses. In: *Journal of Labor Economics*, Vol. 13, No. 2, p. 246–288, URL <http://www.jstor.org/stable/2535104>.

Dustmann, Christian; Glitz, Albrecht (2011): Chapter 4 - Migration and Education. In: Hanushek, Eric A.; Machin, Stephen; Woessmann, Ludger (Eds.) *Handbook of The Economics of Education*, Vol. 4 of *Handbook of the Economics of Education*, Elsevier, p. 327 – 439, URL <http://www.sciencedirect.com/science/article/pii/B9780444534446000043>.

Dustmann, Christian; Soest, Arthur Van (2002): Language and the Earnings of Immigrants. In: *ILR Review*, Vol. 55, No. 3, p. 473–492, URL <https://doi.org/10.1177/001979390205500305>.

Dustmann, Christian; van Soest, Arthur (2001): Language Fluency and Earnings: Estimation with Misclassified Language Indicators. In: *The Review of Economics and Statistics*, Vol. 83, No. 4, p. 663–674, URL <http://www.jstor.org/stable/3211761>.

Fernández-Huertas Moraga, Jesús (2011): New Evidence on Emigrant Selection. In: *The Review of Economics and Statistics*, Vol. 93, No. 1, p. 72–96, URL [https://doi.org/10.1162/REST\\_a\\_00050](https://doi.org/10.1162/REST_a_00050).

Freedom House (2019): *Freedom in the World*.

German Federal Employment Agency (2018): *Employment and Unemployment Statistics*. German Federal Employment Agency.

Goethe Institute (2018a): *Annual Report 2017/2018*. Goethe Institute.

Goethe Institute (2018b): *Articles of Association*. In: of 21 September 2000 as amended on 3 July 2018, URL [https://www.goethe.de/resources/files/pdf165/vereinssatzung-engl\\_30okt18-korr.pdf](https://www.goethe.de/resources/files/pdf165/vereinssatzung-engl_30okt18-korr.pdf).

Goethe Institute (2016): *Basic Agreement*. URL [http://www.goethe.de/resources/files/pdf165/rahmenvertrag-engl\\_30okt18.pdf](http://www.goethe.de/resources/files/pdf165/rahmenvertrag-engl_30okt18.pdf).

Grogger, Jeffrey; Hanson, Gordon H. (2011): Income maximization and the selection and sorting of international migrants. In: *Journal of Development Economics*, Vol. 95, No. 1, p. 42 – 57, URL <http://www.sciencedirect.com/science/article/pii/S0304387810000647>, symposium on Globalization and Brain Drain.

Hanson, Gordon H. (2006): Illegal Migration from Mexico to the United States. In: *Journal of Economic Literature*, Vol. 44, No. 4, p. 869–924, URL <http://www.aeaweb.org/articles?id=10.1257/jel.44.4.869>.

Huber, Matthias; Übelmesser, Silke (2019): Presence of language-learning opportunities and migration. In: CESifo Group Munich, , No. No. 7569.

Imai, Kosuke; Keele, Luke; Yamamoto, Teppei (2010): Identification, Inference and Sensitivity Analysis for Causal Mediation Effects. In: *Statistical Science*, Vol. 25, No. 1, p. 51–71.

Isphording, Ingo E.; Otten, Sebastian (2014): Linguistic barriers in the destination language acquisition of immigrants. In: *Journal of Economic Behavior & Organization*, Vol. 105, p. 30–50.

McKenzie, David; Rapoport, Hillel (2010): Self-Selection Patterns in Mexico-U.S. Migration: The Role of Migration Networks. In: *The Review of Economics and Statistics*, Vol. 92, No. 4, p. 811–821, URL [https://doi.org/10.1162/REST\\_a\\_00032](https://doi.org/10.1162/REST_a_00032).

Melitz, Jacques; Toubal, Farid (2014): Native language, spoken language, translation and trade. In: *Journal of International Economics*, Vol. 93, No. 2, p. 351–363.

Munshi, Kaivan (2003): Networks in the Modern Economy: Mexican Migrants in the U. S. Labor Market\*. In: *The Quarterly Journal of Economics*, Vol. 118, No. 2, p. 549–599, URL <http://dx.doi.org/10.1162/003355303321675455>.

Ramos, Fernando (1992): Out-Migration and Return Migration of Puerto Ricans. University of Chicago Press, p. 49–66, URL <http://www.nber.org/chapters/c6905>.

Roy, A. D. (1951): Some Thoughts on the Distribution of Earnings. In: *Oxford Economic Papers*, Vol. 3, No. 2, p. 135–146, URL <http://www.jstor.org/stable/2662082>.

Schuss, Eric (2018): The Impact of Language Skills on Immigrants' Labor Market Integration: A Brief Revision With a New Approach. In: *The BE Journal of Economic Analysis & Policy*, Vol. 18, No. 4.

UCDP Conflict Encyclopedia (2018): Uppsala Conflict Data Program. URL [www.ucdp.uu.se](http://www.ucdp.uu.se).

Uebelmesser, Silke; Huber, Matthias; Weingarten, Severin (2018): The German Language Worldwide: a New Data Set on Language Learning. In: *CESifo Economic Studies*, Vol. 64, No. 1, p. 103–121, URL <http://dx.doi.org/10.1093/cesifo/ify007>.

Van den Berg, Gerard J (1994): The effects of changes of the job offer arrival rate on the duration of unemployment. In: *Journal of Labor Economics*, Vol. 12, No. 3, p. 478–498.

Van Der Slik, F. W.; Van Hout, R. W.; Schepens, J. J. (2015): The gender gap in second language acquisition: Gender differences in the acquisition of Dutch among immigrants from 88 countries with 49 mother tongues. In: PloS one, Vol. 10, No. 11.

Wissenschaftliche Dienste des Deutschen Bundestages (2006): Das Goethe-Institut: Institutionelle Strukturen und Einbindung in die Auswärtige Kulturpolitik der Bundesrepublik Deutschland. In: Ausarbeitung WD 10 - 043/06.

Worldbank, The (2018): World Development Indicators. Worldbank.

Zschirnt, Eva; Ruedin, Didier (2016): Ethnic discrimination in hiring decisions: a meta-analysis of correspondence tests 1990–2015. In: Journal of Ethnic and Migration Studies, Vol. 42, No. 7, p. 1115–1134.

## Appendix

**Proof of proposition 1.** In the absence of Goethe Institutes, the wage profile faced by potential migrants is given by  $A = \mu_1 + \delta_1 s - e^{\mu_\pi - \gamma_1 s}$ . The effect of Goethe Institutes is to shift the wage profile, which is given by  $B = \mu_1 + \delta_1 s - e^{\mu_\pi - \gamma_1 s(1 + \gamma_2 GI)}$ .

The change in migration incentives,  $\Delta$ , is given by the difference between  $A$  and  $B$ :

$$\Delta = B - A \quad (\text{A1})$$

$$= \mu_1 + \delta_1 s - e^{\mu_\pi - \gamma_1 s(1 + \gamma_2 GI)} - \mu_1 - \delta_1 s + e^{\mu_\pi - \gamma_1 s} \quad (\text{A2})$$

$$= e^{\mu_\pi - \gamma_1 s} \left( 1 - \frac{1}{e^{s\gamma_1\gamma_2 GI}} \right) > 0 \quad (\text{A3})$$

The absence of Goethe Institutes raises migration incentives for any  $s > 0$ ; migration incentives remain unchanged for  $s = 0$ .

**Proof of proposition 2.** The presence of Goethe Institutes raises migration incentives more for persons with high levels of education if  $\frac{\partial \Delta}{\partial s} > 0$ . We have:

$$\frac{\partial \Delta}{\partial s} = -\gamma_1 e^{\mu_\pi - \gamma_1 s} \left( 1 - \frac{1}{e^{s\gamma_1\gamma_2 GI}} \right) + \frac{e^{\mu_\pi - \gamma_1 s} (\gamma_1 \gamma_2 GI)}{e^{s\gamma_1\gamma_2 GI}}. \quad (\text{A4})$$

Because  $-\gamma_1 e^{\mu_\pi - \gamma_1 s} \left( 1 - \frac{1}{e^{s\gamma_1\gamma_2 GI}} \right) < 0$  and  $\frac{e^{\mu_\pi - \gamma_1 s} (\gamma_1 \gamma_2 GI)}{e^{s\gamma_1\gamma_2 GI}} > 0$ ,

$$\frac{\partial \Delta}{\partial s} > 0 \Leftrightarrow -\gamma_1 e^{\mu_\pi - \gamma_1 s} \left( 1 - \frac{1}{e^{s\gamma_1\gamma_2 GI}} \right) < \frac{e^{\mu_\pi - \gamma_1 s} (\gamma_1 \gamma_2 GI)}{e^{s\gamma_1\gamma_2 GI}} \quad (\text{A5})$$

$$\frac{1}{e^{s\gamma_1\gamma_2 GI}} - 1 < \frac{\gamma_2 GI}{e^{s\gamma_1\gamma_2 GI}} \quad (\text{A6})$$

This is always true given that for any  $s > 0$ ,  $\frac{1}{e^{s\gamma_1\gamma_2 GI}} - 1 < 0$  and  $\frac{\gamma_2 GI}{e^{s\gamma_1\gamma_2 GI}} > 0$ .

# Imprint

## **IAB-Discussion Paper IAB 14|2019**

### **Publication date**

8 July 2019

### **Editorial address**

Institute for Employment Research  
of the Federal Employment Agency  
Regensburger Straße 104  
90478 Nürnberg  
Germany

### **All rights reserved**

Reproduction and distribution in any form, also in parts, requires the permission of IAB Nuremberg

### **Download of this Discussion Paper**

<http://doku.iab.de/discussionpapers/2019/dp1419.pdf>

### **All publications in the series “IAB-Discussion Paper“ can be downloaded from**

<https://www.iab.de/en/publikationen/discussionpaper.aspx>

### **Website**

[www.iab.de](http://www.iab.de)

### **ISSN**

2195-2663

---

### **For further inquiries contact the author**

Philipp Jaschke

Telefon 0911 179-2574

E-Mail [Philipp.Jaschke@iab.de](mailto:Philipp.Jaschke@iab.de)

Sekou Keita

Telefon 0911 179-1384

E-Mail [Sekou.Keita@iab.de](mailto:Sekou.Keita@iab.de)