Eric Schuss[1]

The Impact of Language Skills on Immigrants' Labor Market Integration: A Brief Revision With a New Approach

[1] Faculty of Business Administration and Economics, Chair of Public Economics, University of Duisburg-Essen, Weststadttürme Berliner Platz 6-8, 45127 Essen, Germany, E-mail: Eric.Schuss@uni-due.de

Abstract:
I examine the impact of language skills on immigrants' labor market performance by applying a new approach, which allows to estimate wage benefits attributed to initial language skills at arrival. By exploiting unique data, I isolate the endogenous part of current German skills and instrument current command by German proficiency measured retrospectively at the point in time of migration. This approach tackles the problem that labor market effects from current language skills are at risk to reflect merely the sum of a successful residence in Germany and only display growth effects. I find that a good command of German increases labor market income by 47.0% for males, while no significant language effects are detected for females. Further analyses illustrate that differences in language effects by gender can be attributed to selection into different occupations and part-time employments and that language operates complementary and enables cross-border transferability of human capital.

Keywords: transferability of human capital, complementarity of language skills, immigrants, labor market performance

JEL classification: C26, J24, J31, J61

DOI: 10.1515/bejeap-2017-0280

1 Introduction and Prior Research

Does the command of the host country's official language improve causally the labor market integration of immigrants? From the point of view of policymakers, language is one key to a successful integration. In France, a compulsory language test for newly arrived non-Union citizens was introduced on the first of January 2007, which is likely to be followed by language training if the test score was below a certain requirement (Lochmann, Rapoport, and Speciale 2017). The German Law of In-Migration[1] introduced a legal claim on the attendance in a German language course for Union citizens and an obligatory attendance for non-Union citizens. But is there also theoretical and empirical evidence that justifies the promotion of immigrants' language proficiency related to a large amount of public funds?

From the perspective of Economics, language is a main input factor to the production of human capital. Referring to Chiswick and Miller (2002, 2003), Dustmann and Glitz (2011) expose that proficiency in the host country's official language improves cross-border transferability of human capital and labor market experience collected before migration. This complementary function of language skills is empirically confirmed by several studies. At first, Chiswick and Miller (2003) find that one additional year of education raises wages by 5% if the immigrant speaks at least one official language of Canada at home. However, no education effect is found for immigrants who cannot conduct a conversation in English or French. Bleakley and Chin (2004) find that language mainly impacts on wages via increased years of schooling, so that this mediating channel is more important than the direct effects of language skills on labor market outcomes. Berman, Lang, and Siniver (2003) show that Hebrew language proficiency reduces wage differentials (compared to native speakers) of programmers and technicians in Israel, while this does not hold for construction workers and gas station attendants. These results are confirmed by Sanroma, Ramos, and Simon (2008, 2009); Lang and Siniver (2009) and Di Paolo and Raymond (2012).

Next to the high economic relevance of this research question, the sophisticated econometric challenge on how to identify the causal effect of language on labor market outcomes has attracted a large number of studies. Besides studies that focus on measurement errors in capturing language skills (Dustmann and Fabbri 2003;

Eric Schuss is the corresponding author.

Dustmann and van Soest 2001, 2004), several economists tackle the problem of reversed causality and endogenous selection. An established approach is exploiting empirical evidence on language acquisition from the areas of economics and psychobiology, which state that arriving in the host country before the "critical age of language acquisition" results in distinct advantages in learning the host country's official language (Bleakley and Chin 2004). These authors focus on childhood migrants arriving in the United States before the age of 18, and instrument English speaking ability by age at arrival interacted with a dummy for non-English-speaking countries. In the main specification, age at arrival is replaced by another dummy, which indicates whether someone has arrived in the United States before the critical age of language acquisition of 12 or not. Bleakley and Chin (2010), Di Paolo and Raymond (2012), and Miranda and Zhu (2013a, 2013b), and Budria and Swedberg (2015) apply a very similar approach. An important result from previous research is that the ascertained effects by instrumental variable (IV) estimation exceed the coefficients by using ordinary least squares (OLS). Referring to Yao and van Ours (2015), this indicates that upward biases by unobserved heterogeneity and reversed causality are dominated by downward biases from measurement errors. This impressive and recent study based in the Netherlands by Yao and van Ours (2015) instruments Dutch language skills by interacting age at arrival and whether the immigrant spoke other languages than Dutch during childhood. They find a wage disadvantage of language deficiency of 47.9% for women (conditional on employment), while no significant effects were detected for males.[2] By covering all migrant groups with a mean age at arrival of almost 20 years, they overcome the disadvantage in the studies of Bleakley and Chin (2004, 2010) where results are only valid for childhood migrants.

However, if childhood migrants and economic migrants are compared, age at arrival and interaction with information on spoken language during childhood as the instruments are faced with the issue that instrumented language skills do not only capture language skills but different exposures to ethnic networks, different education systems and different migration reasons as well. At that point, my paper inserts and extends the current state of research by several contributions. First, unique data on migrants living in Germany provide retrospective information on socioeconomic status and labor market performance in the country of origin before migration to Germany. Such information is often summarized under unobserved heterogeneities in Migration Economics (Borjas 1994; Chiswick and Miller 1995; Dustmann and van Soest 2002), which, however, likely determine labor market performance in Germany as well as efforts in learning German. Second, this survey dataset also includes a large set of information on initial conditions of migrants at the point in time of migration to Germany such as residency status at arrival and initial German language skills. This latter fact is exploited to examine how initial language skills affect the labor market performance in the long run. By applying an IV approach, I isolate the endogenous part of current command of German language and instrument current command by German proficiency measured retrospectively at the point in time of migration to Germany. This approach subtracts impacts between migration and the current point in time on current language skills and labor market performance and provides empirical evidence that the impact of language on labor market outcomes are not only accumulated growth effects. An initial command of the German language increases the probability of a good current command by 27.2 percentage points. Moreover, I find that German language skills increase males' labor market income marginally by 47.0%, whereas language skills also raise full-time employment probability of males unconditional on employment. Income effects for females are not detected, while only a weakly significant effect on the employment probability is observable. Further examinations illustrate that differences in language effects by gender can be attributed to selection into different occupations and part-time employments. This IV approach also manages to isolate the effect of language from effects of ethnic networks, remittances and migration behavior, which are all parts of the same simultaneous decision function in naive OLS estimations. Third, I show that applying age at arrival and interactions with speaking other languages than German during childhood as the instruments results in a rather large language effect of 111.9%. Thus, these instruments likely overestimate the actual effect of language because this approach cannot isolate the language effect from the effects of ethnic networks, remittances and migration plans. Fourth, robustness checks provide empirical evidence for Germany that using the instrument language skills at arrival demonstrates the complementary function of language proficiency for transferring human capital over international borders.

The rest of the paper is organized as follows. In Section 2, I outline the IV approach as the identification strategy and in Section 3, data and descriptive statistics form the center of attention. Section 4 presents the main results and examines the complementarity of language skills. After Section 5 checks robustness and tackles further economic questions, Section 6 provides concluding remarks and implications.

2 Identification Strategy and Methodology

A couple of issues impede causal interpretation of the effect of language skills on some labor market outcome y_{it}. Besides the problem of reversed causality, measurement errors in self-assessed language skills, unobserved characteristics, and selection produces endogeneity in language skills. For instance, literature on migration and outmigration behavior illustrates that language determines not only labor market integration but also choices of the migrants' residency and that location choices are affected by unobservables such as talent in learning languages and efforts to integrate. Thus, a simple OLS regression of current command of the host country's official language on earnings or employment status y_{it} may just reflect accumulated growth effects that can be attributed to a large set of possible factors, such as endogenous residency choices, efforts to learn German and linkages to ethnic networks, that happened after arrival in the host country. Thus, the supposition that labor market effects of language may only be growth effects accumulated over time also arises from the fact that a large set of impacts after migration determine the command of the host country's official language, so that current skills may be only the sum of the previous success or failure of integration. This paper focuses on this issue of endogeneity and provides a new IV approach to estimate wage returns attributed to initial language skills at arrival. The problem of simultaneity is illustrated in eq. (1), which sources in the circumstance that efforts to learn the host country's language, the decision whether to link to ethnic networks or not, migration plans and indicators of return migration such as remittances are simultaneously determined and part of the same simultaneous decision function. Explanations and a theoretical model by Dustmann and Glitz (2011) support this hypothesis, which shows that the incentive to invest in host country-specific human capital affects simultaneously language skills, tendencies to get connected to ethnic networks and intentions to return or to stay in the host country.

$$y_{it} = \underbrace{\begin{pmatrix} \text{Language skills} \\ \text{Ethnic networks} \\ \text{Migration plans} \\ \text{Remittances} \end{pmatrix}_{it}}_{\text{simultaneous decision function}} \beta + X_{it}\theta + \varepsilon_{it} \tag{1}$$

I apply a two-stage least-squares IV (2SLS IV) approach by instrumenting current German language skills L_{it} of individual i in year t by language skills at arrival L_{it_0} $(t = 0)$. Equation (2) displays the first stage in the IV setting, which furthermore controls the impacts of socioeconomic variables X_{it}, pre-migration characteristics M_i and country of origin-fixed effects δ_k, whereas k identifies different countries and includes unobserved heterogeneities α_i and an idiosyncratic error term u_{it}.

$$L_{it} = L_{it_0}\pi_1 + X_{it}\pi_2 + M_i\pi_3 + \delta_k + \alpha_i + u_{it} \tag{2}$$

In the second stage, I use that part of current German language skills that is explained by initial German skills at arrival $\hat{L}_{it}$ and estimate the language effect β_1 on earnings and employment status y_{it}:

$$y_{it} = \hat{L}_{it}\beta_1 + X_{it}\beta_2 + M_i\beta_3 + \delta_k + \alpha_i + \varepsilon_{it} \tag{3}$$

Theoretical considerations and empirical evidence in Section 4 justify that language skills at arrival are highly positively correlated with language skills during residence in Germany. Furthermore, this approach isolates the endogenous part of current command of German language that can be traced back to impacts and events that happened between migration and the current point in time. Skills of the host country's language at arrival are temporarily observed before any realizations of labor market performance in the host country. Thus, I only use that part of current German language skills that can be explained by initial German skills and thereby tackle the problem that labor market effects of current language skills are at risk to reflect merely the sum of a successful or unsuccessful residence in Germany. Furthermore, empirical results in Section 4 show that arriving in Germany with acquired German language skills is not correlated with a different tendency to get connected to ethnic networks, to send remittances or to have different migration plans, so that this econometric approach is capable of isolating the language effect from other parts of the simultaneous decision function. After a detailed discussion of econometric concerns that impede causal interpretation of language effects, it should be accentuated that my identification strategy does not aim to estimate the causal effect of language on labor market performance. Instead, I focus on my research question and show that current and initial German language skills are highly correlated with each other and that good initial skills at arrival explain a high share of explanatory power of current labor market success. It should be kept in mind that also language skills measured retrospectively at arrival are endogenous determined by reasons of migration and endogenous residency choices.

3 Data and Descriptive Statistics

I use data from the IAB–SOEP Migration Sample for the period 2013–2015 (Wagner et al. 2008; Brücker et al. 2014), which is carried out jointly by the German Institute for Employment Research (Institut für Arbeitsmarkt- und Berufsforschung, IAB) and the German Institute for Economic Research (Deutsches Institut für Wirtschaftsforschung, DIW), which also runs the German Socio-Economic Panel (SOEP). Information on self-assessed proficiency of the German language is collected in three disciplines, namely in speaking, reading and writing, and is distinguished by five different categories: very good, good, intermediate, poor and very poor (see Table 7 in the appendix). Taking account of all three disciplines supplies a more complete and more informative image of language skills, for instance because each discipline is required differently, depending on the kind of industry and on the occupational sector, in which the individual is employed. I define German language skills L_{it} as a binary variable that equals 1 if individual i commands at least one of the three disciplines in a good manner, while the dummy is 0 if he or she does not command any discipline in an at least good manner. The same information is available for German language skills at arrival L_{it_0}, which is constructed analogously in the same way. After excluding persons without any statement on language skills and after excluding circular migrants, the final sample only includes immigrants, which have immigrated to Germany for the very first time, so that a sample of 2,481 male and 2,806 female first-generation immigrants remains.

Table 1: Descriptive statistics by current German language skills.

	Males			Females		
	Good command	No command	Mean diff.	Good command	No command	Mean difference
Socioeconomic variables						
Age	37.4	42.5	5.1***	36.6	41.3	4.7***
Years of residence	14.4	9.6	−4.8***	12.5	10.0	−2.5***
Age at arrival	23.0	33.0	10.0***	24.1	31.4	7.2***
Relationship (in %)	55.3	52.0	−3.3*	52.7	56.9	4.3**
German citizenship (in %)	41.5	19.4	−22.1***	38.8	15.8	−23.0***
Remittances (in %)	15.0	15.2	0.2	13.8	11.0	−2.8**
Informal way of job search (in %)	49.7	57.4	7.7***	42.6	41.2	-1.4
Member of EU (in %)	34.1	40.1	5.9***	37.9	36.5	-1.4
Education (in %)						
High education, ISCED 5–6	23.4	14.0	−9.4***	27.7	15.1	−12.6***
Middle education, ISCED 3–4	46.5	48.0	1.5	46.3	37.2	−9.1***
Low education, ISCED 1–2	26.2	33.7	7.5***	22.0	41.1	19.1***
Graduation in Germany	28.8	2.3	−26.5***	23.0	2.9	−20.1***
Pre-migration characteristics (in %):						
Employed in the home country	51.2	79.6	28.4***	48.1	50.6	2.5
Employee without managerial functions	36.4	61.2	24.8***	36.7	41.6	4.9***
Employee with managerial functions	9.9	11.3	1.4	8.9	6.2	−2.8***
Self-employed	4.7	7.0	2.3***	2.4	2.6	0.2
Relationship in the home country	17.9	35.8	17.9***	22.4	33.6	11.2***
Residency status (in %):						
Current residency status:						
No status	51.5	43.1	8.4***	48.7	38.1	−10.5***
Temporary working permission	9.8	15.3	5.6***	11.2	19.1	7.9***
Permanent right of residency	35.2	33.8	-1.5	37.3	36.1	-1.2
Blue Card	1.2	2.1	0.9**	0.6	1.2	0.7**
Tolerance permit and others	1.3	2.6	1.3***	1.3	3.0	1.7***
Residency status at arrival:						
Family member	29.5	20.3	9.2***	42.8	50.6	7.7***
Asylum seeker	14.3	17.0	2.8**	8.4	12.4	4.0***
Job search	12.9	22.3	9.4***	9.6	13.0	3.4***
Ethnic German	17.0	13.4	−3.6***	18.2	10.0	−8.2***
With job commitment	13.2	21.1	8.0***	6.3	5.7	-0.6
Student and apprentice	12.9	5.2	−7.6***	14.6	8.0	−6.6***

No status	16.9	13.7	-3.2^{**}	18.2	10.2	-8.0^{***}
Observations	2,532	1,051	3,583	3,075	1,070	4,145
Individuals	1,789	865	2,481	2,129	862	2,806

Notes: $^{*}p < 10\%$, $^{**}p < 5\%$, $^{***} < 1\%$. Two-sided t tests check whether means differ significantly to each other, depending on whether individuals display currently good language skills in at least one discipline ($L_{it} = 1$) or not ($L_{it} = 0$).
Source: IAB–SOEP Migration Sample, own illustration.

Table 1 presents the descriptive statistics on regressors used with respect to the current German language skills. Statistics clarify that males and females without any language proficiency are older, arrived in Germany later in their life, display shorter durations of residence and are less likely to be highly educated. Large differences are also visible by examining ethnic networks and residency status. The binary variable informal way of job search indicates that migrants with no language proficiency at all more likely found their first job in Germany via an informal way of job search such as friends, relatives or acquaintances. Migrants without any language skills are more likely to have only temporary working permission and arrived more likely as job seekers, asylum seekers or with a job commitment. The sample comprises mostly economic migrants with a mean age at arrival of 26 years from outside the European Union (64.1% for males and 62.4% for females). The data provide retrospective information on the residency status at arrival and on employment status and occupational positions in the country of origin in the very last year before migration to Germany. The inclusion of these pre-migration characteristics in the regression models largely reduces the danger of biases through unobserved heterogeneities, which potentially affect both current labor market performance and language skills (Willis and Rosen 1979; Borjas 1994). Unobserved heterogeneities possibly enhance the positive link and lead to an upward bias and overestimation of the impact of language skills, because highly motivated migrants put more effort into their working life and in learning German (Borjas 1994; Dustmann and van Soest 2002; Yao and van Ours 2015). On the other hand, downward biases by unobserved heterogeneity are also conceivable, which might strengthen the downward bias by measurement errors.[3] Thus, the direction of this issue is not completely clear; however, controlling for it is undoubtedly necessary (Chiswick and Miller 1995; Dustmann and van Soest 2002). So far, research has tended to approximate the employment biography in the country of origin, for instance by the socioeconomic status of the individual's partner and by the parental education (Dustmann and van Soest 2002). Thus, having this retrospective information is quite important.

Figure 1 illustrates a first look at (annual) labor market income distributions, which is my main indicator of labor market performance, and indicates a positive relationship between language proficiency and integration. A good command in language results in a significant descriptive annual wage advantage of € 3,752.4 for males and € 5,178.4 for females. Differences in wage distributions are also valid if language skills at arrival are considered, whereas 19.1% of the whole sample arrive in Germany with a good command of German and the rest without any language proficiency. Additional descriptive statistics in Table 8 in the appendix provide insights into the sample composition by gender with respect to occupations and employment status. In Section 5, I analyze reasons for different language effects by gender in more detail. However, a quite different composition with respect to occupations and employment status is already apparent in Table 8.

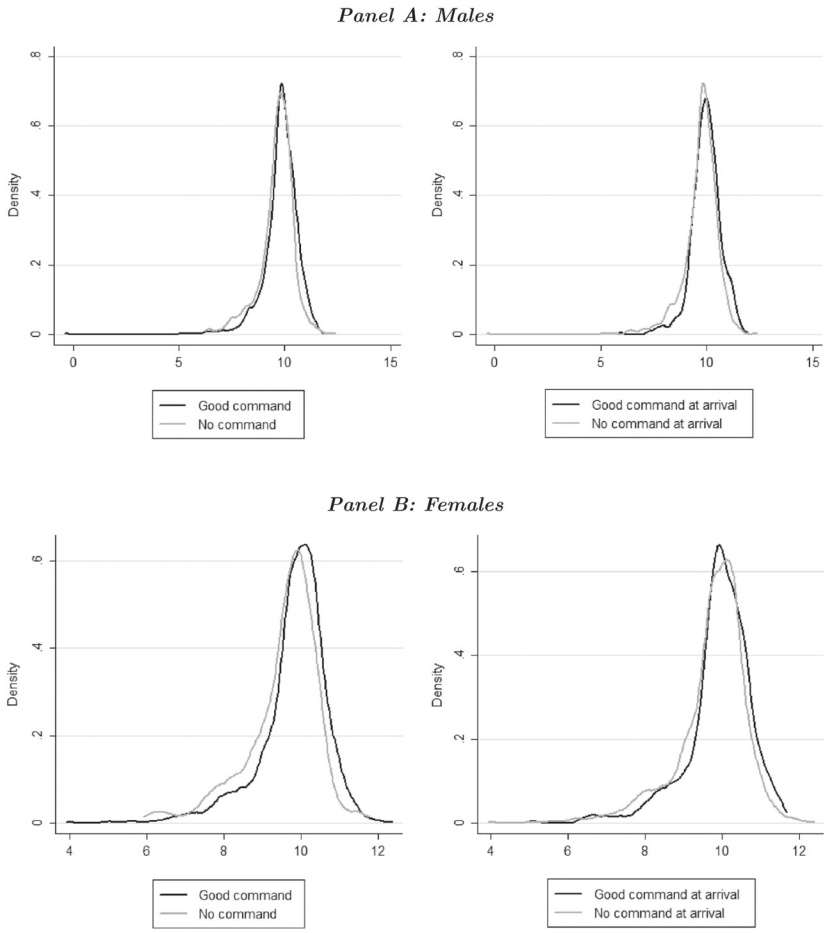

Figure 1: Distributions of logarithmized labor market income by current and initial German language skills at arrival. Source: IAB-SOEP Migration Sample, own illustration

4 Empirical Evidence

4.1 Main Results

Table 2 presents determinants of logarithmized labor market income per year from OLS and 2SLS IV estimations of eq. (3) with random effects conditional on employment for males. OLS estimations of β_1 in eq. (3) show that migrants with a good command of at least one discipline out of speaking, reading or writing earn an 8.8% higher labor market income. The given numbers for the coefficient of determination R^2 and for χ^2 clarify the importance of including further controls and particularly controlling on pre-migration characteristics and the residency status at arrival. This highlights that having information on these retrospective information improves strongly the reliability of the presented estimations. Coming to IV estimations, first, estimation of eq. (2) shows a high relevance of the instrument used, which is displayed in the lower panel. In Model (6), a good command of German at arrival increases the probability of having currently good language skills by 27.2 percentage points, whereas in each model, the test of weak instruments displays a high F statistic of at least 79.6. This fulfills the requirements to a relevant instrument (Angrist and Pischke 2009). The upper panel of Table 2 shows finally that good language skills traced back to initial skills at arrival increase marginally labor market income by 47.0%, so that naive OLS estimations underestimate the true language effect. This indicates that the downward bias produced by measurement errors of self-assessed language skills dominate possible upward biases by reversed causality and unobserved heterogeneities (see also Bleakley and Chin 2004; Yao and van Ours 2015). This is consistent with Yao and van Ours (2015), who find also a domination of measurement errors and an effect of language deficiency of about 47.9% for females, a result quite similar in magnitude to mine, however, for the other sex. Table 9 displays OLS and IV results for females without any significant effects after adding controls. This is why proceeding analyses are restricted to males.

Table 2: The effect of German language skills on logarithmized labor market income for males (conditional on employment).

	OLS			2SLS IV		
	(1)	(2)	(3)	(4)	(5)	(6)
Language skills	0.1047***	0.0773**	0.0881***	0.8108***	0.4810***	0.4698**
	(0.0303)	(0.0307)	(0.0306)	(0.1949)	(0.1835)	(0.1838)
Socioeconomic variables:						
Age		0.0037	0.0011		0.0078**	0.0045
		(0.0024)	(0.0025)		(0.0030)	0.0030
Years of residence		0.0087***	0.0133***		0.0031	0.0084**
		(0.0031)	(0.0032)		(0.0041)	(0.0040)
Relationship		-0.0677	-0.0486		−0.0844*	-0.0620
		(0.0431)	(0.0452)		(0.0445)	(0.0463)
German citizenship		0.0768	0.0880		0.0205	0.0426
		(0.0571)	(0.0597)		(0.0653)	(0.0665)
Remittances		0.1553***	0.1495***		0.1639***	0.1583***
		(0.0305)	(0.0303)		(0.0319)	(0.0317)
Informal way of job search		-0.0328	-0.0457		-0.0140	-0.0294
		(0.0368)	(0.0361)		(0.0384)	(0.0374)
Member of EU		0.1936***	0.0300		0.2135***	0.0450
		(0.0693)	(0.0708)		(0.0711)	(0.0720)
Education:						
High education, ISCED 5-6		0.3315***	0.3191***		0.2645***	0.2654***
		(0.0448)	(0.0458)		(0.0550)	(0.0534)
Low education, ISCED 1-2		−0.2040***	−0.2006***		−0.1746***	−0.1736***
		(0.0425)	(0.0420)		(0.0451)	(0.0442)
Graduation in Germany		-0.0222	0.0159		-0.0651	-0.0113
		(0.0624)	(0.0662)		(0.0662)	(0.0682)
First stage:						
Language skills at arrival				0.3010***	0.2779***	0.2724***
				(0.0339)	(0.0307)	(0.0304)
F statistic				79.6***	81.9***	80.2***
Additional control variables:						
Country of origin		Yes	Yes		Yes	Yes
Current residency status		Yes	Yes		Yes	Yes
Pre-migration characteristics			Yes			Yes
Residency status at arrival			Yes			Yes
R^2	0.0105	0.1783	0.2171	0.0105	0.1570	0.1983
χ^2	12.0	405.7***	514.2***	17.3***	392.2***	500.2***
Number of observations	2,626	2,626	2,626	2,626	2,626	2,626
Individuals	1,861	1,861	1,861	1,861	1,861	1,861

Notes: $^{*}p < 10\%$, $^{**}p < 5\%$, $^{***}p < 1\%$; standard errors in parentheses. Table 2 considers males and uses logarithmized labor market income per year conditional on employment as the dependent variable. Models (1)–(3) apply random effects GLS and Models (4)–(6) apply two-stage least-squares instrumental variable (2SLS IV) estimations, where current command of the German language is instrumented by German language skills at arrival. The models are extended stepwise with socioeconomic and educational covariates, country of origin-fixed effects, current residency status, pre-migration characteristics and initial residency status at arrival. A list on used covariates can be found in Table 1.
Source: IAB–SOEP Migration Sample, own illustration.

Note for completeness that the language effect in specification (6) increases in magnitude if ethnic Germans or childhood migrants who arrived in Germany before the age of 18 are excluded (see Table 10 in the appendix). Furthermore, results are robust to the inclusion of immigration year-fixed effects, in order to determine whether results are distorted because different migration waves differ substantially with respect to socioeconomic status, migration reasons and contemporary historical background. The same holds for adding area-fixed effects on the level of federate states and for the application of estimation between effects and pooled least-squares estimations.[4]

Further regressors in the main model (6) confirm further results from the literature. Consistent with Akresh (2006, 2008), integration improves with duration of residence, while ethnic networks, captured by an informal way of job search, and German citizenship are not significantly correlated with labor market income. In accordance with Clark and Drinkwater (2002), education is an important determinant of labor market success, whereas Bleakley and Chin Bleakley and Chin (2004) and Dustmann and Glitz (2011) accentuate the importance of language skills as a main input factor to the production function of human capital that ensures the transferability of human capital to another country. Bleakley and Chin (2004) find with further investigations that this mediating channel is far more important than the direct effects of language on labor market outcomes. Other studies found that only the labor market performance of migrants with a high level of education or with highly demanding occupations benefit from language skills (e.g. Berman, Lang, and Siniver 2003; Chiswick and Miller 2003; Di Paolo and Raymond 2012). This is left to examine in the following section.

4.2 The Complementarity of Language Proficiency

To ensure that language only impacts labor market outcomes via improving the transferability of collected human capital, the exogeneity of my instruments needs to be analyzed. In detail, first, it must be ensured that applying the instrument language skills at arrival isolates the labor market effect of language from other parts of the simultaneous decision function, displayed in eq. (1). Afterwards, I can examine the complementarity of language skills to the production function of human capital in a second step. These two steps are analyzed stepwise in this section.

Table 3: Variables of the simultaneous decision function and differences with respect to language skills at arrival.

	by (language skills at arrival)			
	Mean difference	**Standard error**	**Observations**	**Individuals**
Ethnic networks:				
Informal way of job search	0.0319	0.0233	3,583	2,481
Support with immigration	0.0177	0.0231	3,583	2,481
Migration plans:				
Intention to return	-0.0192	0.0238	2,360	2,360
Intention to stay	0.0242	0.0250	2,360	2,360
Remittances	-0.0195	0.0167	3,583	2,481
Pre-migration characteristics:				
Employed in the home country (in %)	-0.0087	0.0229	3,583	2,481
Employee without managerial functions (in %)	0.0357	0.0231	3,583	2,481
Employee with (in %) managerial functions (in %)	-0.0476^{***}	0.0141	3,583	2,481
Self-employed at home (in %)	0.0010	0.0105	3,583	2,481
Relationship in the home country (in %)	0.0384^{*}	0.0196	3,583	2,481

Notes: $^{*}p < 10\%$, $^{**}p < 5\%$, $^{***}p < 1\%$. Two-sided t tests check whether means differ significantly to each other, depending on whether males display a good command of German at arrival in at least one discipline ($L_{it_0} = 1$) or not ($L_{it_0} = 0$). The two dummies, irrespective of whether individuals have the intention to return and to stay, are only available for two years, namely 2013 and 2015.
Source: IAB–SOEP Migration Sample, own illustration.

First, Table 3 tests the exogeneity of my instrument and lists parts of the simultaneous decision function in eq. (1). The table presents indicators of ethnic networks, migration plans including remittances and pre-migration characteristics and tests whether migrants with German language skills at arrival vary in their likelihood to get linked to ethnic networks, have different migration plans and display a different retrospective labor market performance in the country of origin, compared to migrants without any language proficiency at arrival. To provide a more complete image, I add support with immigration as one further indicator of ethnic networks, which indicates whether a person received help from family members, friends or acquaintances with his or her moving to Germany or not. This binary variable was not included in my basic set of regressors, used in Table 1 and Table 2. Moreover, I add two further dummies to Table 3, namely whether the surveyed immigrant currently plans to return to the country of origin and whether the migrant plans to stay in Germany forever or not. Both variables are only available for two waves, namely for 2013 and 2015. Simple t tests clarify that means in ethnic networks, migration plans, remittances and pre-migration characteristics do not differ significantly

with respect to language skills at arrival. The only issue is that migrants with language skills at arrival were more likely employees with managerial functions in the country of origin. As the only variable, however, this is negligible, because the general employment status does not differ with respect to language skills. Finally, it can be concluded that language skills at arrival is an exogenous instrument which is not correlated with a different tendency to get connected to ethnic networks, to send remittances and to have different migration plans or a different past employment biography.

Table 4: Stratified estimations: Language skills as a complementary input factor for the production of human capital.

	Second stage	**First stage**		
	Language skills	**Language skills at arrival**	***F* statistic**	**Observations [Individuals]**
Education:				
High and middle education	0.5371**	0.2564***	55.7***	1,953
	(0.2208)	(0.0344)		[1,383]
High education	0.4565	0.1883***	15.4***	618
	(0.5326)	(0.0480)		[442]
Middle education	0.5049**	0.2926***	40.5***	1,304
	(0.2303)	(0.0460)		[921]
Low education	0.2313	0.2991***	18.8***	673
	(0.3706)	(0.0686)		[486]
Years of schooling:				
Yrs of schooling ≤ 8	0.5190***	0.3043***	59.6***	1,867
	(0.1916)	(0.0395)		[1,342]
Yrs of schooling ≤ 9	0.5422**	0.2926***	44.6***	1,522
	(0.2216)	(0.0438)		[1,104]
Yrs of schooling ≤ 10	0.5326**	0.2876***	41.8***	1,362
	(0.2352)	(0.0445)		[990]
Yrs of schooling ≥ 8	0.4023	0.3548***	19.0***	450
	(0.3263)	(0.0813)		[106]
Yrs of schooling ≥ 9	0.3417	0.3371***	20.1***	610
	(0.2999)	(0.0753)		[427]
Yrs of schooling ≥ 10	0.4957**	0.3496***	41.8***	966
	(0.2143)	(0.0544)		[672]
Prestige:				
Prestige, SIOPS > 30	0.5510***	0.2424***	45.7***	1,938
	(0.2396)	(0.0358)		[1,391]
Prestige, SIOPS ≥ 30	0.2780	0.3756***	36.5***	688
	(0.2590)	(0.0622)		[497]

Notes: $^{*}p < 10\%$, $^{**}p < 5\%$, $^{***}p < 1\%$; standard errors in round brackets. Table 4 presents 2SLS IV estimations for males conditional on employment with logarithmized labor market with income as the dependent variable. Each presented estimation of this table uses language proficiency at arrival as the instrument and controls on the same set of covariates as Model (6) in Table 2. A list on used covariates can be found in Table 1. Estimations are restricted to certain education groups and are stratified with respect to years of schooling and prestige scores. Prestige is measured by the Standard International Prestige Scale (SIOPS) developed by Ganzeboom and Treiman (1996, 2003).
Source: IAB–SOEP Migration Sample, own illustration.

After a detailed examination of Table 3, it can be concluded that the presented IV results in Table 2 seem to be reliable for capturing only the effect of language skills and isolating the language effects from other parts of the simultaneous decision function, displayed in eq. (1). Next is the requirement for analysis of whether the new instrument, language proficiency at arrival, ensures demonstration of the complementary function of language skills for the transferability of human capital to another country. This is examined by following the approach of Berman, Lang, and Siniver (2003), Chiswick and Miller (2003) and Di Paolo and Raymond (2012), who suggest stratified estimations with respect to indicators of human capital. Table 4 presents results from IV regressions with language skills at arrival. With respect to the International Standard Classification of Education (ISCED), German language skills clearly increase labor market income for the pooled group of high- and middle-qualified groups (ISCED 3–6), whereas no significant language effect can be detected for low-educated migrants. However, separate stratifications for high- and middle-educated immigrants show that language is only beneficial for middle-educated immigrants. The finding that language improves labor market outcomes of immigrants with middle and high human capital but not for immigrants with a low endowment of human capital is confirmed by an alternative indicator of human capital, namely years of schooling. Language

is only beneficial for immigrants with at least 8, 9 and 10 years of schooling, while no significant language effect can be detected for immigrants with maximal 8 and maximal 9 years of schooling. This confirms results from an application of an endogenous switching model on migrants living in Catalonia (Di Paolo and Raymond 2012). A proficiency in the Catalan language is followed by a significant wage advantage of 22.2% for migrants with more than 8 years of schooling, while no significant effect is shown for migrants with 8 years of schooling or less. My results are also in line with Chiswick and Miller (2003), who find that wage benefits from an additional year of education increase to 5% for immigrants who speak at least one official language of Canada at home, while no education effect can be found for those immigrants who cannot conduct a conversation in English or French. I introduce the Standard International Occupational Prestige Scale (SIOPS) of Ganzeboom and Treiman (1996, 2003), which captures occupational quality of the current working position and reflects required qualifications and education for certain occupations. Referring to Fasani (2015), I define jobs with a prestige score of at least 30 as low-quality occupations. Results in Table 4 demonstrate that a significant language effect on labor market income is detected for immigrants currently employed in middle- and high-demanding occupations. On the contrary, labor market performance of migrants with a prestige score no more than 30 does not benefit from proficiency of the German language. This confirms the results of Berman, Lang, and Siniver (2003), who find that Hebrew language proficiency reduces the wage gap between migrants and natives for programmers and technicians in Israel, while this does not hold for construction workers and gas station attendants. On the whole, after having shown that language skills at arrival are not correlated with a current linkage to ethnic networks, migration plans, and pre-migration characteristics, education and human capital were significantly identified as the mediating channel between language skills and labor market outcomes.

5 Robustness Checks and Further Analyses

This section has a focus on further econometric and economic issues. First of all, I compare estimates with my new recommended instrument to approaches, which use age at arrival and different interaction terms between age at arrival and information on spoken language during childhood as the instruments. As outlined in the econometric Section 2, an instrument should focus on isolating the effect of language from the effects of ethnic networks, migration plans and remittances on labor market outcomes. These are the other parts of a simultaneous decision function, which are determined simultaneously as parts of the same decision process. However, in a sample which consists of both childhood migrants, who arrived in Germany before the age of 18, and economic migrants, the instrument age at arrival probably does not only capture language, but also labor market effects from ethnic networks and different education systems and attendance. Table 5 illustrates this presumption for males. In Models (1) and (2), I apply the instrument language skills at arrival L_{it_0} as the instrument, where Model (1) equals Model (6) of the main Table 2 and Model (2) only considers migrants who arrived before the age of 40. These estimations are compared to Models (3)–(5), in which I apply the instrument from Bleakley and Chin (2004, 2010) , and Yao and van Ours (2015). In Model (3), current language skills are instrumented by an interaction term between age at arrival and whether a person spoke a different language than German during childhood.[5] Model (4) additionally includes age at arrival as a second instrument and model (5) focuses analogously to Model (2) on migrants arriving in Germany before the age of 40. Estimations of the first stage illustrate that these alternative instruments are highly relevant as well, whereas the Sargan test indicates that using two instruments does not lead to a problem of overidentification. Comparisons of the baseline model (1) to the specifications (3) and (4) indicate that age at arrival results in a rather large labor market effect of 110.6% and 110.2%, which presumably is not isolated from other parts of the simultaneous decision function in eq. (1). However, if the sample is restricted to migrants arriving before the age of 40, coefficients in both models (2) and (4) decrease, which indicates that the reliability of instruments highly depends on the composition of the sample with respect to age at arrival and migration groups.[6]

Table 5: IV estimations with alternative instruments.

	Baseline model		**Alternative instruments**		
	(1)	**(2) Age at arrival ≥ 40**	**(3)**	**(4)**	**(5) Age at arrival ≥ 40**
Language skills	0.4698**	0.3628*	1.1059***	1.1019***	0.6288***
	(0.1816)	(0.2159)	(0.2482)	(0.2392)	(0.2362)
First stage:					
Language skills at arrival	0.2724***	0.2425***			

	(0.0304)	(0.0315)			
Age at arrival × spoken other languages during childhood			-0.0100^{***} (0.0012)	-0.0586^{***} (0.0022)	-0.0055^{**} (0.0024)
Age at arrival				-0.0062^{**} (0.0027)	-0.0083^{***} (0.0031)
F statistic	80.2^{***}	58.7^{***}	67.1^{***}	36.4^{***}	31.6^{***}
Sargan test of overidentification				0.007	0.079
Additional controls	Yes	Yes	Yes	Yes	Yes
R^2	0.1983	0.2019	0.1295	0.1299	0.1786
χ^2	500.2^{***}	441.8^{***}	516.8^{***}	519.1^{***}	518.2^{***}
Number of observations	2,626	2,362	2,626	2,626	2,362
Individuals	1,861	1,665	1,861	1,861	1,665

Notes: $^{*}p < 10\%$, $^{**}p < 5\%$, $^{***}p < 1\%$; standard errors in parentheses. Table 5 presents 2SLS IV estimations for males conditional on employment with logarithmized labor market income per year as the dependent variable. In Models (1) and (2), German language commands at arrival are used as the instrument, whereas Model (2) only considers immigrants, which arrived before the age of 40. In Model (3), current language skills are instrumented by an interaction term between age at arrival and whether one spoke a different language than German during childhood. The last named dummy is approximated by an interaction term of whether the person was born in a country that has another official language than German or not and whether the person arrived in Germany before the age of 18 or not. Model (4) includes age at arrival as the second instrument and model (5) focuses on migrants, which arrived in Germany before an age of 40. Each specification controls the same set of covariates as Model (6) in Table 2. A list on used covariates can be found in Table 1. Source: IAB–SOEP Migration Sample, own illustration.

The second concern tackled in this section is the question why income effects from language skills are only found for males, but not for females. A first claim may be that this is due to using annual labor market income as y_{it}, so that persons working part-time or that are marginal employed suffer under the choice of this outcome variable. However, using hourly wage conditional on employment as an alternative outcome still results in a highly significant effect in robust magnitude for males and insignificance for females.[7] A second approach to tackle this claim is to estimate the effect of German language skills on labor market entry. Therefore, Models (1) and (2) in Table 6 apply the presented IV approach on the probability to be employed and to be full-time employed. Again a highly positive full-time employment effect from language skills is detectable for males, while the language effect on employment probabilities for females is only weakly significant. Thus, language effects are also found to be stronger for males at the extensive margin. The comparison of the extensive to intensive margin also illustrates that the sample conditional on employment is positive selected in terms of language skills and further socioeconomic characteristics (see also Edo, Jacquement, and Yannelis 2017). This should be kept in mind when deriving implications from my results.

Table 6: IV estimation on labor market entry and stratified IV analyses by occupational sectors.

	Panel A: Males							
	(1) Employment	**(2) Full-time employment**	**(3) Blue collar**		**(5) White collar**		**(7) Public sector**	
			Pr(blue collar)	**Income**	**Pr(white collar)**	**Income**	**Pr(public sector)**	**Income**
Language skills	0.0764	0.1799**	−0.1818*	0.4755**	0.0949	0.3457	0.0301	2.1193
	(0.0791)	(0.0875)	(0.1040)	(0.2038)	(0.1046)	(0.3084)	(0.0594)	(2.8041)
First stage:								
Language skills at arrival	0.2794***	0.2808***	0.2680***	0.3355***	0.2681***	0.2182***	0.2700***	0.0912
	(0.0230)	(0.0243)	(0.0243)	(0.045)	(0.0244)	(0.0404)	(0.0261)	(0.0941)
F statistic	149.26***	136.5***	121.2***	55.1***	120.3***	29.15***	106.6***	1.0
Additional controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.1479	0.1956	0.2108	0.1888	0.1697	0.2481	0.0388	0.1521
χ^2	454.7***	653.2***	534.9***	257.6***	409.1***	270.6***	82.5***	39.8
Number of observations	3,583	3,583	2,626	1,272	2,626	1,056	2,626	186
Individuals	2,481	2,481	1,861	959	1,861	812	1,861	146
	Panel B: Females							
	(1) Employment	**(2) Full-time employment**	**(3) Blue collar**		**(4) White collar**		**(5) Public sector**	
			Pr(blue collar)	**Income**	**Pr(white collar)**	**Income**	**Pr(public sector)**	**Income**
Language skills	0.1905*	-0.0171	−0.4327***	0.3086	0.3752**	0.2893	-0.0797	-0.1633
	(0.0985)	(0.0890)	(0.1384)	(0.3404)	(0.1501)	(0.4548)	(0.1201)	(0.8831)
First stage:								
Language skills at arrival	0.2189***	0.2212***	0.1847***	0.3039***	0.1859***	0.1300***	0.1889***	0.1215***
	(0.0185)	(0.0199)	(0.0210)	(0.0624)	(0.0214)	(0.0255)	(0.0232)	(0.0423)
F statistic	140.2***	124.4***	75.8***	23.0***	73.2***	26.0***	63.6***	8.3***
Additional controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

R^2	0.1819	0.1176	0.1960	0.1550	0.1379	0.2023	0.0239	0.3237
χ^2	700.9^{***}	414.3^{***}	440.8^{***}	141.9^{***}	270.7^{***}	250.9^{***}	61.0^{***}	146.8^{***}
Number of observations	4,145	4,145	2,359	796	2,359	1,279	2,359	400
Individuals	2,806	2,806	1,690	631	1,690	968	1,690	308

Notes: $^{*}p < 10\%$, $^{**} < 5\%$, $^{***}p < 1\%$; standard errors in parentheses Models (1) and (2) replace labor market income as the outcome variable by binary information on employment and full-time employment, so that probabilities for labor market entry are estimated. In these estimations, the full sample is used unconditional on employment. In Models (3), (5) and (7), the probability to be employed in the blue collar, white collar or in the public sector is estimated by using again the sample conditional on employment. Models (4), (6) and (8) again apply logarithmized labor market income per year as the outcome variable and stratify estimations on the respectively occupational sector. Each estimation is based on IV 2SLS estimations, applies initial German language skills as the instrument and controls on the full set of covariates (see Table 1).
Source: IAB–SOEP Migration Sample, own illustration.

A second potential reason to explain different language effects in terms of gender is selection into occupations, in which return of language skills differ. To check this claim, Table 6 first of all estimates the effect of German language skills on the probability to be employed in blue-collar, white-collar jobs or in the public service in Models (3), (5) and (7). IV results illustrate that a good command of the German language decreases the probability to work in the blue-collar sector for both females and males. Particularly, language skills reduce this probability for females by 43.3 percentage points, while language increases the probability to be employed in a white-collar occupation for females by 37.5 percentage points. On the contrary, language does not significantly increase the probability to be employed in white-collar occupations for males. Then, in a second step, IV estimations of language skills on logarithmized labor market income are stratified on the respective occupational sector in Models (4), (6) and (8). Those analyses clearly illustrate that an income benefit from language skills is only detectable for males that are employed in blue-collar occupations.

A closer look on German proficiency reveals that 89.2% of females occupied in the white-collar sector displays a good command of German. This fact and the absence of any language effect for persons, occupied in white-collar occupations, demonstrate that a reliable degree of German proficiency is a basic requirement to be hired in a white-collar job and reasons at the same time why proficiency of the German language does not result in higher income returns.[8]

These considerations come together with sexual selection into occupations, illustrated by the fact that the share of white-collar workers is larger for females by 14.0 percentage points compared to males (see Table 8 in the appendix). To put in nutshell, this selection into occupations in terms of gender and a low variation in German proficiency in the sector of white-collar jobs are a further reason that helps to explain the transmission channel between language and labor market performance and why language effects differ between females and males.

6 Conclusion

I examine the impact of language skills on migrants' labor market performance in Germany on a sample mainly consisting of economic migrants with a mean age at arrival of 26 years from outside the European Union. I apply a new approach to estimate which share of current labor market success is attributed to initial German proficiency at arrival and exploit unique data, which provides several contributions: First, unique data on migrants living in Germany from the IAB–SOEP Migration Sample provides retrospective information on socioeconomic status and labor market performance in the country of origin before migration to Germany. Such information is often unobserved (Borjas 1994; Chiswick and Miller 1995; Dustmann and van Soest 2002) and likely determine labor market performance in Germany as well as efforts in learning German. Although talent in learning languages and individual disutility from working is still unobservable, OLS and IV estimation demonstrate that explanatory power of my estimated models is strongly improved by implementing these variables into my models. Second, having retrospective information on initial German language skills at arrival is exploited to analyze how initial German proficiency affects current labor market performance. This approach isolates the endogenous part of current command of German language that can be traced back to impacts and events that happened between migration and the current point in time. By instrumenting current language proficiency by past proficiency at arrival, only that part of current German language skills is used that can be explained by initial skills. Thereby, the problem is tackled that labor market effects of current language skills are at risk to mere reflect the sum of a successful or unsuccessful residence in Germany.

This identification strategy does not identify the causal effect of language on labor market performance because also language skills measured retrospectively at arrival are endogenous, and are for instance determined by endogenous residency choices. Despite this limitation and although that applying language skills at arrival as the instrument is not applicable with fixed effects estimation, this paper nevertheless contributes to the literature by showing that current and initial German language skills are highly correlated with each other and that German proficiency traced back to initial proficiency significantly improves labor market performance. I find that an initial command of the German language increases the probability of a good current command by about 27.2 percentage points. Moreover, I estimate that German language skills increase labor market income marginally by 47.0% for employed males and increases full-time employment by 18.0 percentage points, while no language effects are detectable for females. Further examinations illustrate that differences in language effects by gender can be attributed to selection into different occupations and that due to this selection, income returns from language differs highly with respect to the choice of the occupational sector.

Further analyses provide the third contribution and confirm empirical evidence from Berman, Lang, and Siniver (2003), Bleakley and Chin (2004), Chiswick and Miller (2003), and Di Paolo and Raymond (2012) and others, who identify education as a main mediating channel between language and labor market outcomes. Stratified estimations demonstrate that language skills only increase earnings of immigrants with at least 8

years of schooling and earnings of immigrants employed in occupations with prestige scores larger than 30. However, a special focus on education groups show that language is mainly important for middle-educated migrants, whereas language turns into insignificance again for migrants with tertiary education degrees including academics (high education, ISCED 5–6). This demonstrates that language skills are a complementary good that enables transferability of human capital to another country. Contrary to prior research is the finding that the German language is mostly important for occupations with middle requirements, while language is not that important in academic workplaces such as international concerns and universities and in occupations with low requirements, e.g. in the manufacturing industry. Fourth, this new approach supplements original work by Bleakley and Chin (2004), which instrumented language by interactions between age at arrival and countries of origin for a sample of children migrants arriving in the United States before the age of 18. Robustness checks show that point estimates depend highly on the sample composition in terms of age. Comparisons with this method demonstrate that my approach results in a more conservative point estimate, considering a sample that consists mainly of economic migrants.

Appendix

Table 7: Different disciplines of German language command over 2013–2015 (in %).

	Males				Females			
	2013	**2014**	**2015**	**2013–2015**	**2013**	**2014**	**2015**	**2013–2015**
Speaking								
Very good	31.8	33.2	13.0	28.3	32.1	33.2	14.0	29.0
Good	37.0	34.3	32.0	35.0	38.6	33.6	30.0	35.2
Intermediate	25.5	25.3	34.5	27.3	22.2	24.9	35.8	25.7
Poor	5.4	6.6	18.9	8.7	6.6	7.1	17.6	8.9
Very poor	0.4	0.7	1.7	0.8	0.5	1.3	2.6	1.2
∑	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Observations	1,603	1,223	757	3,583	1,905	1,440	800	4,145
Reading								
Very good	31.8	32.4	14.7	28.4	34.6	34.9	17.6	31.4
Good	36.7	35.6	33.0	35.6	39.3	37.4	37.3	38.3
Intermediate	22.3	23.6	29.7	24.3	17.4	17.9	27.8	19.6
Poor	7.1	6.4	17.4	9.0	6.0	6.0	12.8	7.3
Very poor	2.1	2.0	5.2	2.7	2.6	3.7	4.6	3.4
∑	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Observations	1,603	1,223	757	3,583	1,905	1,440	800	4,145
Writing								
Very good	27.8	26.7	11.4	24.0	28.9	30.2	13.8	26.4
Good	30.1	29.2	25.9	28.9	35.4	32.4	29.5	33.2
Intermediate	25.6	27.0	29.7	26.9	22.1	22.4	32.8	24.3
Poor	12.6	13.0	24.4	15.2	10.2	10.6	16.9	11.6
Very poor	4.0	4.1	8.6	5.0	3.5	4.4	7.1	4.5
∑	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Observations	1,603	1,223	757	3,583	1,905	1,440	800	4,145

Source: IAB–SOEP Migration Sample, own illustration.

Table 8: Sample composition by labor market performance and countries of origin.

	Males			Females			
	Good command	**No command**	**Mean diff.**	**Good command**	**No command**	**Mean diff.**	**Mean diff. by sex**

Labor market performance:							
Intensive margin:							
Labor market income (in € per year)	23,621.8	19,869.4	$-3,752.4^{***}$	24,649.3	19,476.3	$-5,178.4^{***}$	$1,000.5^{**}$
Hours weekly worked	40.8	40.3	-0.5	29.7	25.6	-4.1^{***}	-11.7^{***}
Employed (in %)	79.6	67.4	-12.2^{***}	63.6	44.7	-19.0^{***}	-17.3^{***}
Full-time employed	66.2	57.2	-9.0^{***}	26.6	15.7	10.9^{***}	-39.7^{***}
Part-time employed	8.6	9.8	1.2	33.2	28.1	-5.0^{***}	22.9^{***}
Not employed	20.4	32.6	12.2^{***}	36.4	55.3	19.0^{***}	17.3^{***}
Occupational sectors (in %):							
Blue collar	42.5	65.5	23.0^{***}	27.4	60.0	32.6^{***}	-14.7^{***}
White collar	44.6	27.8	-16.8^{***}	60.0	30.1	-29.9^{***}	14.0^{***}
Public sector	8.0	4.6	-3.4^{***}	18.6	10.3	-8.3^{***}	9.9^{***}
Self-employed	4.7	6.5	1.8^{*}	2.5	3.5	1.0	-2.4^{***}
Countries of origin (in %):							
Greece	3.5	3.5	3.5	2.6	3.2	2.7	-0.7^{*}
Italy	5.2	5.3	5.2	3.2	2.9	3.1	-2.1^{***}
Spain	1.5	1.5	1.5	2.3	1.1	2.0^{**}	0.5
Turkey	9.9	11.0	10.3	6.1	14.9	8.3^{***}	-2.0^{**}
(Former) Yugoslavia	13.0	6.9	11.2^{***}	9.9	9.8	9.8	-1.4^{**}
Countries of EU enlargement 2004	11.4	14.7	12.4^{***}	15.3	14.7	15.1	2.7^{***}
Countries of EU enlargement 2007	9.2	12.4	10.1^{***}	11.7	12.4	11.9	1.8^{**}
Russia and (former) USSR	29.0	24.7	27.7^{***}	34.8	19.4	30.8^{***}	3.1^{***}
East, Southeast and remaining Asia	2.4	3.0	2.6	3.0	5.4	3.6^{***}	1.0^{**}
Northern Africa	2.4	2.9	2.6	1.1	2.1	1.4^{**}	-1.2^{***}
Africa	2.1	2.0	2.1	1.7	2.6	1.9^{**}	-0.2
Arabic countries	6.6	8.4	7.1^{*}	4.7	8.8	5.7^{***}	-1.4^{**}
Observations	2,532	1,051	3,583	3,075	1,070	4,145	7,728
Individuals	1,789	865	2,481	2,129	862	2,806	5,275

Notes: Two-sided *t* tests: *significant at 10%, **significant at 5%, ***significant at 1%
Source: IAB–SOEP Migration Sample, own calculations.

Table 9: Baseline estimations for females: The effect of German language skills on logarithmized labor market income (conditional on employment).

	OLS			**2SLS IV**		
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Language skills	0.1284^{***}	0.0547	0.0487	0.7514^{***}	0.3892	0.2465
	(0.0398)	(0.0399)	(0.0396)	(0.2432)	(0.2714)	(0.2634)
Socioeconomic variables:						
Age		-0.0086^{***}	-0.0102^{***}		-0.0056	-0.0084^{**}
		(0.0025)	(0.0027)		(0.0035)	(0.0036)
Years of residence		0.0045	0.0099^{**}		0.0010	0.0077
		(0.0039)	(0.0039)		(0.0047)	(0.0048)
Relationship		0.3003^{***}	0.3370^{***}		0.3037^{***}	0.3404^{***}
		(0.0447)	(0.0479)		(0.0453)	(0.0482)
German citizenship		0.0347	0.0307		-0.0187	0.0008
		(0.0621)	(0.0635)		(0.0770)	(0.0761)
Remittances		0.1887^{***}	0.1796^{***}		0.1849^{***}	0.1805^{***}
		(0.0370)	(0.0368)		(0.0385)	(0.0380)
Informal way of job search		0.0048	-0.0066		0.0042	-0.0070
		(0.0413)	(0.0407)		(0.0416)	(0.0405)
Member of EU		0.2713^{***}	0.1385^{*}		0.2523^{***}	0.1291
		(0.0809)	(0.0817)		(0.0828)	(0.0822)

Education:						
High education, ISCED 5-6		0.3850***	0.3471***		0.3449***	0.3256***
		(0.0489)	(0.0491)		(0.0589)	(0.0568)
Low education, ISCED 1-2		−0.2224***	−0.1940***		−0.1846***	−0.1740***
		(0.0514)	(0.0510)		(0.0611)	(0.0592)
Graduation in Germany		-0.0918	-0.0210		-0.1110	-0.0319
		(0.0696)	(0.0710)		(0.0718)	(0.0721)
First stage:						
Language skills at arrival				0.2321***	0.1888***	0.1916***
				(0.0273)	(0.0257)	(0.0258)
F statistic				65.2***	51.4***	52.1***
Additional control variables:						
Country of origin		Yes	Yes		Yes	Yes
Current residency status		Yes	Yes		Yes	Yes
Pre-migration characteristics			Yes			Yes
Residency status at arrival			Yes			Yes
R^2	0.0143	0.1875	0.2119	0.0143	0.1756	0.2088
χ^2	10.4***	378.0***	2,441,86.4***	9.6***	374.1***	469.5***
Number of observations	2,359	2,359	2,359	2,359	2,359	2,359
Individuals	1690	1690	1690	1690	1690	1690

Notes: $^{*}p < 10\%$, $^{**}p < 5\%$, $^{***}p < 1\%$; standard errors in parentheses. Table 2 considers females and uses logarithmized labor market income per year conditional on employment as the dependent variable. Models (1)–(3) apply random effects GLS and Models (4)–(6) apply two-stage least-squares instrumental variable (2SLS IV) estimations, where current command of the German language is instrumented by German language skills at arrival. The models are extended stepwise with socioeconomic and educational covariates, country of origin-fixed effects, current residency status, pre-migration characteristics and initial residency status at arrival. A list on used covariates can be found in Table 1.
Source: IAB–SOEP Migration Sample, own illustration.

Table 10: Minor robustness checks.

	Males		**Females**	
	(1) Ethnic Germans excluded	**(2) Age at arrival ≥ 18**	**(3) Ethnic Germans excluded**	**(4) Age at arrival ≥ 18**
Language skills	0.4723**	0.5361***	0.5210	0.4484*
	(0.2184)	(0.3553)	(0.2573)	(0.2573)
First stage:				
Language skills at arrival	0.2699***	0.3112***	0.1727***	0.2127***
	(0.0339)	(0.0375)	(0.0303)	(0.0303)
F statistic	63.5***	69.0***	29.9***	48.7***
Additional controls	Yes	Yes	Yes	Yes
R^2	0.2118	0.2011	0.2109	0.2048
χ^2	450.5***	401.2***	421.3***	393.0***
Number of observations	2,181	2,030	1,919	1,925
Individuals	1,573	1,458	1,407	1,382

Notes: $^{*}p < 10\%$, $^{**}p < 5\%$, $^{***}p < 1\%$; standard errors in parentheses. Table 10 presents 2SLS IV estimations for males and females conditional on employment with logarithmized labor market income per year as the dependent variable. Each estimation controls on the full set of covariates summarized in Table 1. Model (1) and (3) excludes ethnic Germans from estimations and Model (2) and (4) excludes childhood immigrants, which arrived before the age of 18 in Germany.
Source: IAB–SOEP Migration Sample, own illustration.

Notes

1 "The Law about Governance and Control of Immigration and about Regulation of Residence and Integration of Union Citizens and Foreigns" (*Gesetz zur Steuerung und Begrenzung der Zuwanderung und zur Regelung des Aufenthalts und der Integration von Unionsbürgern und Ausländern, Zuwanderungsgesetz*)
2 This is consistent with Lindley (2002), who identifies language fluency as one determinant of the wage gap between migrants and natives, while for males ethnicity is more important in decomposing this wage gap.
3 One thesis regarding this effect is offered by Dustmann and van Soest (2002). "This is the case if foregone earnings of individuals who engage in language education increase with their unobserved ability." (See Willis and Rosen 1979.)
4 Furthermore, results are robust when bootstrap and jackknife heteroskedastic standard errors are applied. These estimation results are available from the author upon request. Note that due to the time-invariant instrument language skills at arrival and a large number of time-invariant regressors, fixed effects estimation is not applicable here.
5 The information regarding whether a person spoke another language than German during childhood is not directly observable. I approximate this dummy by an interaction term between whether the person was born in a country with another official language than German or not, and whether the person arrived in Germany before the age of 18 or not. However, countries having German or similar languages as the official language are only Austria, Switzerland and Benelux countries, which covers only a very low number of individuals in my sample. Thus, exogeneous similarity in German and foreign language cannot be exploited forward.
6 Note that the language effect instrumented by language skills at arrival increases again to 43.4% and becomes again highly significant ($p < 0.05$) if only migrants are considered with an age at arrival between 18 and 40.
7 Presentation of these estimations are omitted and results are available from the author upon request.
8 A similar argumentation would be plausible for jobs in the civil sectors. However, a very low number of observations impede reliability of Models (7)–(8).

References

Akresh, I. R. 2006. "Occupational Mobility among Legal Immigrants to the United States." *International Migration Review* 40 (4): 854–885.
Akresh, I. R. 2008. "Occupational Trajectories of Legal US Immigrants: Downgrading and Recovery." *Population and Development Review* 34 (3): 435–456.
Angrist, J. D., and J. S. Pischke. 2009. *Mostly Harmless Econometrics*. Princeton University Press.
Berman, E., K. Lang, and E. Siniver. 2003. "Language-Skill Complementarity: Returns to Immigrant Language Acquisition." *Labour Economics* 10: 265–290.
Bleakley, H., and A. Chin. 2004. "Language Skills and Earnings: Evidence From Childhood Immigrants." *Review of Economics and Statistics* 86 (2): 481–496.
Bleakley, H., and Chin, A. 2010. "Age at Arrival, English Proficiency, and Social Assimilation Among US Immigrants." *American Economic Review: Applied Economics* 2 (1): 165–192.
Borjas, G. J. 1994. "Self-Selection and the Earnings of Immigrants." *American Economic Review* 77 (4): 531–553.
Borjas, G. J. 1995. "Ethnicity, Neighborhoods, and Human-Capital Externalities." *American Economic Review* 85 (3): 365–390.
Brücker, H., M. Kroh, S. Bartsch, J. Goebel, S. Kühne, E. Liebau, P. Trübswetter, I. Tucci, and J. Schupp. 2014. "The New IAB-SOEP Migration Sample: An Introduction Into the Methodology and the Contents." *SOEP Survey Paper 216*, Series C, DIW Berlin.
Budria, S., and P. Swedberg. 2015. "The Impact of Language Proficiency on Immigrants Earnings." *Revista de Economa Aplicada* 67: 63–91.
Chiswick, B. R., and P. W. Miller. 1995. "The Endogeneity between Language and Earinings: International Analyses." *Journal of Labor Economics* 13 (2): 246–288.
Chiswick, B. R., and P. W. Miller. 2002. "Immigrant Earnings: Language Skills, Linguistic Concentration and the Business Cycle." *Journal of Population Economics* 15 (1): 31–57.
Chiswick, B. R., and P. W. Miller. 2003. "The Complementarity of Language and Other Human Capital: Immigrant Earnings in Canada." *Economics of Education Review* 22 (5): 469–480.
Clark, K., and S. Drinkwater. 2002. "Enclaves, Neighbourhood Effects and Employment Outcomes: Ethnic Minorities in England and Wales." *Journal of Population Economics* 15: 5–29.
Di Paolo, A., and J. L. Raymond. 2012. "Language Knowledge and Earnings in Catalonia." *Journal of Applied Economics* 15 (1): 89–118.
Dustmann, C., and F. Fabbri. 2003. "Language Proficiency and Labour Market Performance of Immigrants in the UK." *The Economic Journal* 113: 695–717.
Dustmann, C., and A. Glitz. 2011. "Migration and Education." In: *Handbook of the Economics of Education*, vol. 4, Chapter 4, edited by E. A. Hanushek, S. Machin and L. Woessmann, 327–439. North-Holland.
Dustmann, C., and A. van Soest. 2001. "Language Fluency and Earnings: Estimation with Misclassified Language Indicators." *The Review of Economics and Statistics* 83 (4): 663–674.
Dustmann, C., and A. van Soest 2002. "Language and the Earnings of Immigrants." *Industrial and Labor Relations Review* 55 (3): 473–492.
Dustmann, C., and A. van Soest. 2004. "An Analysis of Speaking Fluency of Immigrants Using Ordered Response Models with Classification Errors." *Journal of Business and Economic Statistics* 22 (3): 312–321.
Edo, A., N. Jacquement, and C. Yannelis. 2017. "Language Skills and Homophilous Hiring Discrimination: Evidence From Gender and Racially Differentiated Applications." *Review of Economics of the Household* forthcoming.
Fasani, F. 2015. "Understanding the Role of Immigrants' Legal Status: Evidence from Policy Experiments." *CESifo Economic Studies* 61 (3): 722–763.
Ganzeboom, H. B. G., and D. J. Treiman. 1996. "Internationally Comparable Measures of Occupational Status for the 1988 International Standard Classification of Occupations." *Social Science Research* 25: 201–239.

Ganzeboom, H. B. G., and D. J. Treiman. 2003. "Three Internationally Standardised Measures for Comparative Research on Occupational Status." In: *Advances in Cross-National Comparisons: A European Working Book for Demographic and Socio-Economic Variables*, edited by J. H. P. Hoffmeyer-Zlotnik and C. Wolf, 159–193. New York: Springer US.

Lang, K., and F. Siniver. 2009. "The Return to English n a Non-English Speaking Country: Russian Immigrants and Native Israelis in Israel." *The B.E. Journal of Economic Analysis & Policy* 9 (1): 1–28.

Lindley, J. 2002. "The English Language Fluency and Earnings of Ethnic Minorities in Britain." *Scottish Journal of Political Economy* 49 (4): 467–487.

Lochmann, A., H. Rapoport, and B. Speciale. 2017. "The Effect of Language Training on Immigrants' Economic Integration - Empirical evidence from France." *CESifo Working Paper* No. 6460.

Miranda, A., and Y. Zhu. 2013a. "The Causal Effect of Deficiency at English on Female Immigrants' Labor Market Outcomes in the UK." *IZA Discussion Paper* No. 7841.

Miranda, A., and Y. Zhu. 2013b. "English Deficiency and the Native Immigrant Wage Gap." *Economics Letters* 118 (1): 38–41.

Sanroma, E., R. Ramos, and H. Simon. 2008. "The Portability of Human Capital and Immigrant Assimilation: Evidence for Spain." *IZA Discussion Paper* 3649: 1–36.

Sanroma, E., R. Ramos, and H. Simon. 2009. "Immigrant Wages in the Spanish Labour Market: Does the Origin of Human Capital Matter? *IZA Discussion Paper* 4157: 1–25.

Schuss, E. 2016. "Between Life Cycle Model, Labor Market Integration and Discrimination: An Econometric Analysis of the Determinants of Return Migration." *SOEPpapers on Multidisciplinary Panel Data Research, DIW Berlin* No. 881.

Wagner, G. C., J. Göbel, P. Krause, R. Pischner, and I. Sieber. 2008. "Das Sozio-oekonomische Panel (SOEP): Multidisziplinäres Haushaltspanel und Kohortenstudie für Deutschland - Eine Einführung (für neue Datennutzer) mit einem Ausblick (für erfahrene Anwender)." *Wirtschafts- und Sozialstatistisches Archiv* 2: 301–328.

Willis, R. J., and S. Rosen. 1979. "Education and Self-Selection." *The Journal of Political Economy* 87 (5, part 2): 7–36.

Yao, Y., and J. C. van Ours. 2015. "Language Skills and Labor Market Performance of Immigrants in the Netherlands." *Labour Economics* 34: 76–85.