Review of Income and Wealth
Series 64, Number 3, September 2018
DOI: 10.1111/roiw.12299

LITERACY AND THE MIGRANT–NATIVE WAGE GAP

BY OLIVER HIMMLER

Max Planck Institute, Bonn

AND

ROBERT JÄCKLE*

Technische Hochschule Nürnberg Georg Simon Ohm

Being able to read and write is one of the most important skills in modern economies. Literacy frequently is a prerequisite for employment and its relevance for productivity and wages is magnified by the fact that it is only through literacy that many other skills become usable. More so than for natives, this argument applies to migrants: even those with high levels of human capital acquired in the country of origin often have it rendered worthless by the absence of literacy in the host-country language. Using novel data from a large-scale German adult literacy test ("leo.—Level-One Studie", or "LEO"), we investigate the determinants of literacy and show that migrants have systematically lower language skills than natives. We find that any observed raw employment and wage gaps between natives and migrants can be fully explained by these differences.

JEL Codes: J24, J31, J61

Keywords: literacy, migration, wage gap, discrimination

1. INTRODUCTION

Migrants earn less than natives. This blanket statement describes the so-called "wage gap" between natives and migrants that has been observed for many years in many countries. Earnings differentials are found in the United States (U.S.) (Chiswick, 1978; Borjas, 1985; Sanders and Lessem, 2013), Canada (Ferrer *et al.*, 2006), the United Kingdom (U.K.) (Chiswick, 1980; Bell, 1997; Denny *et al.*, 1997; Miranda and Zhu, 2013), and Germany (Pischke, 1992; Dustmann, 1993; Aldashev *et al.*, 2012; Bartolucci, 2014), among others. Non-trivial differences often remain even after factoring out determinants of wages such as educational attainment (see, *inter alia*, Algan *et al.*, 2010).

It is important to investigate whether such observed inequalities can be attributed to migrants actually being discriminated against by employers or

Note: We are grateful to two anonymous referees for helpful suggestions, and we thank seminar and conference participants at CReAM (UC London), Miami, Munich, and Würzburg, as well as Thomas Cornelissen and Tobias König, for helpful comments and discussions. We also thank Susanne Jäckle for the construction of the language tree dummies. Biljana Meiske and Peter Frenzel provided excellent research assistance.

*Correspondence to: Robert Jäckle, Technische Hochschule Nürnberg Georg Simon Ohm, Department of Business Administration, Bahnhofstrasse 87, 90402 Nürnberg, Germany (robert.jaeckle@th-nuernberg.de).

whether they simply lack competencies or qualifications that are relevant in the labor market, and those are not captured in the data. One important qualification for many jobs is the ability to read and write in the native tongue of the country in which one resides. Migrants likely differ from natives in this respect, and if they indeed show lower levels of literacy in the host-country language, failing to factor in these differences may be responsible for at least part of any observed wage gap.

One reason is that lack of literacy can in general impair productivity; for example, because it keeps an individual from communicating with others, from following work instructions and safety regulations, or from acquiring further human capital and gathering information on, for instance, how to maintain their health. This is true for natives and migrants alike, but in the case of migrants, being able to read and write in the native tongue of a country is complementary to any human capital they may have already acquired in their native country. Not being able to read and write can void educational attainment and occupational qualifications accumulated in the country of origin prior to migrating.

This complementarity of literacy arises, of course, because it is a necessary requirement in order to be able to apply many skills—think of the migrant engineer or physician who cannot communicate with clients, colleagues, or patients because he does not speak the language. The same mechanism applies to lower qualification levels as well, and even those who are qualified for jobs that do not require communicating with others on the job at all may see their skills invalidated if, for example, they cannot find the right job because they are unable to read job offers. While this complementarity theoretically also matters for natives, one is far less likely to observe a native with a high stock of human capital who does not speak the native language than to observe a migrant for whom this is true. In fact, in the data that we use, only 4 percent of natives with a university degree are functionally illiterate in the German language, whereas this applies to 24 percent of migrants. This high prevalence of "mismatches" between education levels and literacy among migrants may lead to the often observed phenomenon of migrants downgrading to jobs below their formal qualification level (Friedberg, 2001; Ozden *et al.*, 2005; Dustmann *et al.*, 2013), and therefore may substantially contribute to wage differentials with equally educated natives.

The assessment of whether language proficiency is responsible for observed wage differences requires a good measure of literacy. Objective measures of literacy are not easy to come by, and this is why many studies use self-reported assessments of how well individuals speak a language (Chiswick, 1991, 1992; Chiswick and Miller, 1995—for a concise survey, see Chiswick and Miller, 2015). Prominent examples include the U.S. census or the German Socioeconomic Panel (SOEP). These assessments are, of course, subject to measurement error, and potentially also to systematic bias, because self-assessments depend on the individual reference group, that is, the local standard of literacy. This is evident in Finnie and Meng (2005), who find that migrants who earn more are more likely to assess their language skills in a positive manner, which tends to overstate the importance of literacy for earnings. They use data that contain both literacy test scores as well as self-assessments, and show that test scores are a superior measure compared to self-assessed literacy.

It is also possible that migrants generally underestimate their language skills—for example, because they compare themselves to those natives who are perfectly literate—which would lead researchers to overestimate the portion of any migrant–native wage gap that is due to literacy. Along these lines, questions about literacy in surveys are often not asked of natives at all. If information on the literacy of natives is not available, by looking at whether the wage gap disappears for migrants who have very good language proficiency one might end up comparing highly proficient migrants with the average skilled native. This is not necessarily informative because it implicitly assumes that every native person has good literacy skills in their native tongue, which is certainly not true, as we will show.

The use of test scores rather than self-assessments reduces problems due to measurement error, one of the main issues that has plagued the literature. When self-assessed literacy scores are used, instrumental variable estimates often produce larger coefficients on literacy than the OLS estimates, and suggest that the downward bias in estimates caused by measurement error in the language variables is large (Dustmann and van Soest, 2001, 2002; Bleakley and Chin, 2004). Having a precise measure of language skills is thus obviously desirable, but in our case an additional requirement is that it should be particularly selective in the lower ranges of literacy. The reason is that—as we will show—a large fraction of migrants do not have a good command of German, and therefore the wage gap will to a large extent be determined by variation in this literacy region.

The data that we use are especially well suited for this purpose and to our knowledge are novel to the economics literature. It stems from the "leo.—Level-One Studie" (hereinafter "LEO"), which was conducted by the University of Hamburg. LEO is the first large-scale German literacy survey which explicitly focuses on the lower end of the skill spectrum, "Level One." Upon its release, LEO gained quite some media attention, mainly because it uncovered the fact that the prevalence of illiteracy in Germany is roughly twice as high as previously thought. Some 8,400 individuals were interviewed, representative of the German population. To give an idea of how LEO compares in terms of difficulty to the International Adult Literacy Survey (IALS), the most well-known literacy test, the lowest IALS level is roughly equivalent to the fifth-lowest LEO level. LEO is less selective at the upper end of the spectrum, but roughly 35 percent of natives and 76 percent of migrants in our data fall into the lower range (below the lowest IALS level), where IALS cannot differentiate but LEO can identify four skill levels ranging from "strict illiteracy" to "below grade school level."

Ability bias in the literacy coefficient is the second problem with which the literature on the general effects of literacy on wages needs to cope. This issue is, of course, not solved by using test scores, and we will not be able to cleanly disentangle the effects of literacy on wages from those of unobserved ability. However, this does not harm our specific analysis, since we focus not on causally estimating the returns to literacy per se, but rather on the question of whether any raw wage gap can be explained by literacy—or any other productivity relevant factor that may be captured by literacy. Obviously, factors such as motivation and ability should be rewarded on the labor market, and with cross-sectional data these skills may be partly reflected in the literacy variable for both migrants and natives.

Fortunately, ability bias is not a big concern for our analysis as long as ability is captured in the literacy of migrants and natives alike. To further support this argument, we show that there are no significant interaction effects between migrant status and literacy, and we are therefore confident that the literacy variable does not measure different things for the two groups of individuals. A third concern is reverse causality. Employment or higher wages might, for example, lead to higher literacy especially in migrants, and thus affect our estimates of the employment and wage gaps. However, we do not find an interaction effect between literacy and migrant status. Assuming that the true returns to literacy are the same for migrants and natives, this could be taken as tentative evidence that reverse causality does not drive our results.

This is the backdrop against which we will attempt to paint an encompassing picture of how language proficiency relates to the performance of migrants on the German labor market, compared to natives. First, we investigate the determinants of language proficiency in the population, and we assess to what extent the literacy skills of migrants are systematically different from those of native Germans. There is a clear expectation that migrants fare on average worse than Germans, and we show that they indeed on average have lower test scores by about one standard deviation. Across both groups, those who are more highly educated tend to have higher literacy skills. For migrants, literacy also improves with time since migration, and those individuals whose native language is more similar to German fare better on the test, although the differences across language groups are not particularly large.

Having established that migrants and natives differ greatly in terms of German literacy, in the next step we look at whether these differences are reflected on the labor market. Specifically, we ask whether migrants are less likely to be employed, and whether any potential employment gap between migrants and Germans can be explained by differences in literacy. The initial differential in employment is roughly 6 percent to the disadvantage of migrants, and considering differences in education cuts this gap in half. Migrants who have spent more time in Germany are also more likely to be employed, as are migrants whose native language is closer to German. The latter is true even conditional on literacy, suggesting that linguistic distance may also be a proxy for factors such as cultural distance, differences in work ethics, missing networks, or informational deficiencies with respect to the host-country labor market. For both natives and migrants, lower literacy levels are associated with significantly lower probabilities of being employed, ranging from 4 to 15 percentage points when compared to individuals who can at least read and write at a fourth-grade level (LEO level >4). Because of the above-mentioned lower literacy skills of migrants, taking into account language proficiency in our estimations further reduces the differential between the groups, to the extent that it explains literally all of the remaining employment gap.

This then leads to the third question we address in this paper: what is the importance of literacy for the earnings of both migrants and natives, and can differences in literacy explain observed wage differences between migrants and natives? The results follow a pattern that is very similar to the one in the employment equations: controlling for education cuts the initial 14 percent wage disadvantage of migrants in half, and on top, literacy is very closely related to wages. Those who are illiterate in the strictest sense command 27 percent lower earnings than those who reach at least LEO level five. As a consequence, when correcting for the fact that migrants have lower proficiency in German, no wage gap remains. These results provide tentative evidence that the raw differences in earnings may not be rooted in discrimination but can actually be explained by observable skills that are relevant for productivity. We show that our findings are robust to the use of different definitions of migrant status, and across a number of subsamples, where the results also exhibit the same pattern.

Within the literature on the earnings of migrants, our paper is most closely related to research on migration and the economics of language. Most of the work in this area is concerned with explicitly estimating the returns to language skills for non-natives. This is in contrast to our goal of explaining the wage gap, which does not rely on clean identification of returns to literacy. Dustmann and Glitz (2011) and Chiswick and Miller (2015) provide excellent surveys of the international evidence, and here we focus on work that has employed German data: Dustmann (1994) uses data from the first wave of the SOEP, which provide self-assessments of migrant language skills, and shows that higher levels of literacy go hand in hand with higher earnings of migrants. Dustmann and van Soest (2001) revisit the topic, and show that the self-assessed language skills from the SOEP are subject to severe misclassification. Using instrumental variables, they show that measurement error leads to a substantial downward bias of the OLS literacy coefficients. Aldashev *et al.* (2012) also rely on SOEP data, and show that increased language skills go hand in hand with increased labor market participation, and that those with a better command of German are more likely to earn higher wages because they are more likely to be employed in white-collar jobs.

Through our estimation of the determinants of literacy, we also share common ground with the emerging literature on the effects of linguistic origin and linguistic distance on the acquisition of the destination language. In line with what we find in our literacy equations, Isphording and Otten (2014) and Isphording (2014) show that migrants with a greater linguistic distance between destination language and their native language are at a disadvantage.

Finally, our paper is also related to the economics literature on cognitive skills. Literacy is often considered to be a cognitive skill, or at least a measure of cognitive skills, and the idea that cognitive skills are one qualification that crucially affects economic outcomes is, of course, not new. At the macro level, countries which have a larger human capital stock at their disposal outperform those whose population lacks basic skills. At the micro level, a staple result is that those with higher cognitive skills earn more. Azariadis and Drazen (1990), Coulombe *et al.* (2004), and Coulombe and Tremblay (2006) show that higher levels of literacy are reflected in economic growth. At the individual level, a number of studies find high returns to literacy (McIntosh and Vignoles, 2001; Green and Riddell, 2003; Vignoles, 2010—for a survey, see Hanushek and Woessmann, 2008).

In addition, the analysis is also linked to the vast literature on discrimination of migrants and ethnic minorities. In a field experiment Kaas and Manger (2012) show that job applicants with German-sounding names are more likely to receive a callback in the application process than those who have a Turkish-sounding name. The effect disappears when the applications include identical reference letters. This is consistent with statistical discrimination: in the absence of information, employers use ethnic origin as a proxy for productivity relevant features of an applicant (such as literacy skills). However, in another field experiment, Sprietsma (2013) finds evidence that identical student essays obtain significantly lower grades and lower secondary school recommendations when they bear a Turkish-sounding rather than a German-sounding name. While our analysis implies that there is no discrimination when employers have to decide between migrants and natives with identical literacy skills, the latter experiment suggests that migrants who attend school in Germany may be discriminated against before they even arrive on the labor market.

The remainder of the paper is structured as follows: Section 2 introduces the data, explains sample adjustments, and gives variable definitions. Section 3 investigates the determinants of literacy and differences between natives and migrants. Sections 4 and 5 are concerned with the employment and wage gaps between migrants and natives, and Section 6 contains extensions and consistency checks. Section 7 concludes.

2. Data and Descriptives

The data used in our analysis are taken from the "leo.—Level-One Studie" (LEO) provided by the University of Hamburg (see Grotlüschen and Riekmann, 2011). LEO is representative of the German population of migrants and natives aged between 18 and 64, whose language skills are sufficient to respond to a German survey interview. The interview as well as the literacy tests are conducted in German only, and accordingly the test measures literacy in the German language.[1] LEO conducts practical reading and writing tests as part of the face-to-face interview, which enables us to cope with the substantial measurement error introduced by the self-reported language proficiency scales that many studies use (Carliner, 1981; Chiswick, 1991—for a comprehensive survey, see Dustmann and Glitz, 2011). These competence tests allow for a categorization of the sample population into five groups, as follows. Respondents who are able to read and write at the *letter level* (= α-level 1) and at the *word level* (= α-level 2) are *strict illiterates.* They can logographically identify single words from graphic features (α-level 1), or they may be able to read or write single words (α-level 2) but not sentences. Those who are at the *sentence level* (= α-level 3) are *functionally illiterate*, that is, they are able to read and write single sentences but fail even with short texts (for sample test items, see Figures A.4 and A.5 in the Appendix, in the Online Supporting Information). Respondents *below grade school level* (= α-level 4) cannot read or write texts at a level that is expected at the end of fourth grade—these people typically avoid reading and writing, even with texts that include only commonly known words. The final group consists of those whose literacy skills are at or above the grade school level (= α-level 5).

By explicitly focusing on the lower end of the literacy scale (= Level One), the LEO study provides a novel dataset and fills a gap in the existing literature.[2]

[1] An individual may therefore be illiterate in German and literate in another language.

[2] Detailed information on how LEO allows us to differentiate these low skill levels and how the LEO items compare with other literacy tests can be found at http://blogs.epb.uni-hamburg.de/leo/.

LEO includes 8,436 observations from across all of the German states (Bundesländer) and consists of two subsamples. The larger sample (7,035 units) is randomly drawn from among the German population. The smaller sample (1,401 observations) is selected from the population of people with secondary education or below—a means to sample more individuals who are not able to read and write sufficiently.[3] Combining the two subsamples generates different selection probabilities for individuals with higher or lower school degrees. Therefore, we apply probability weights in our estimations and when making inference to the population in the descriptive statistics.

We use observations from both subsamples and extract data on the variables described in Tables A.1 and A.2. As recommended by Bilger *et al.* (2012) in the LEO technical manual, we drop 20 observations with obviously invalid information on the literacy variables. We constrain the sample to individuals who belong to the labor force, that is, those who are engaged in full-time or part-time work and individuals who report to be currently unemployed. We further exclude 44 units who cannot be uniquely assigned to one educational level. After these steps, we obtain a final unweighted estimation sample of 5,651 observations, including 568 migrants. Of these, 3,107 individuals (including 291 migrants) are full-time employed, 1,418 (136) are engaged in part-time work, and 1,126 (141) are unemployed.

Taking migrants and natives together, *strict illiteracy* (=α-levels 1 and 2) affects 4.4 percent of the labor force (see Figure A.2), 9.8 percent of the labor force is *functionally illiterate*, and another 25.7 percent of the workforce cannot read or write at a level that is expected at the end of fourth grade (α-level 4). In sum, roughly 40 percent of the labor force only have very limited reading and writing skills at their disposal.

2.1. *Migration, Wage, and Literacy Variables*

The Migration Variable

The literature defines *migrants* in a number of ways. The most exclusive operationalization is to only consider those who do not have citizenship of the respective country. Often, migrants are also defined to mean those who were born abroad—which, in comparison to citizenship, adds individuals who obtained host-country citizenship post migration, and excludes those who were born in the host country but do not have citizenship.[4]

We later also provide robustness checks using the above definitions, but since we are interested in the relation between language proficiency and labor market outcomes—not citizenship effects or born abroad effects per se—we adopt another definition: we label as migrants all individuals who do not report their native tongue to be German (we include in the native group those individuals who grew up bilingually with German and another language).[5] In comparison to the above definitions, this assigns foreign citizens (those born in a foreign

[3]For further details, see Bilger *et al.* (2012).

[4]Unlike many other countries, children born in Germany do not automatically obtain German nationality.

[5]By using this definition, our group of migrants is expected to be subject to Chiswick's (1978, 1979) model of "positive assimilation"—whereas common language would speak for "negative assimilation."

country) who name German as their first language to the migrant group, and adds citizens of the host country (people born in Germany) who do not report German to be their mother tongue to the native group.

Given our definition, 13.6 percent of the labor force and 12.8 percent of the employed qualify as migrants. Using the standard definitions, 9.4 percent (8.7 percent) of the labor force (employed) are foreign citizens and 15.8 percent (14.9 percent) were not born in the host country (see Table A.2).

In order to illustrate that LEO is representative of the German labor market, we compare our numbers with data provided by the German Statistical Office (2011) for 2009—one year before the LEO data were collected.[6]

According to the Statistical Office, in 2009, 15.3 percent of the labor force and 14.4 percent of the employed were born abroad; 7.9 percent (7.3 percent) of the labor force (employed) were reported to be foreign citizens. These numbers very closely match the figures in the LEO data (see Table A.2).

Literacy Variables and Plausible Values

Instead of a single cognitive literacy score, the LEO dataset includes five *plausible values*. These values are random draws from the posterior distribution of a latent variable, given each individual's responses to the test items and a set of background variables in a conditioning discrete choice model. The latter assumes the literacy skills to be normally distributed among the population.[7] When plausible values are used, measurement error in the literacy scores is negligible (see Junker *et al.*, 2012) and the efficiency of population estimates improves. However, as each draw of a plausible value includes a random error component, these values cannot be individually allocated as test scores. Therefore, similar to analyses using multiple imputations (see Rubin, 1987), we run a regression on each plausible value, average the results, and adjust the standard errors for variation between the five estimates.

In addition to the continuous literacy scores, LEO also provides five discrete α-levels (see Figure A.2). To convert the continuous score into literacy levels, LEO defines thresholds, which are anchored in the LEO pretest and earlier literacy studies. A person reaches a certain α-level if they can solve a typical item from the corresponding level of difficulty with a probability of 62 percent (because we have five plausible values, a single individual can be allocated to different α-levels in different draws).

The Wage Variable

The LEO study measures wages as monthly gross income from current employment. As is well known from the literature on survey methodology, asking for information on wages is an intrusive question for many respondents. In the LEO survey, 37 percent of the income data is missing. However, 87 percent of those refusing to quote their salary were willing to classify it within certain ranges ($\leq$ €400, €401–1,000, and $>$ €1,000). For those who only provided a class of

[6]The German Statistical Office provides detailed migration statistics only every four years. We thus compare the LEO survey to official statistics from 2009.

[7]For further details, see Hartig and Riekmann (2012). A practical guide to constructing and applying plausible values can be found in Adams and Wu (2002).

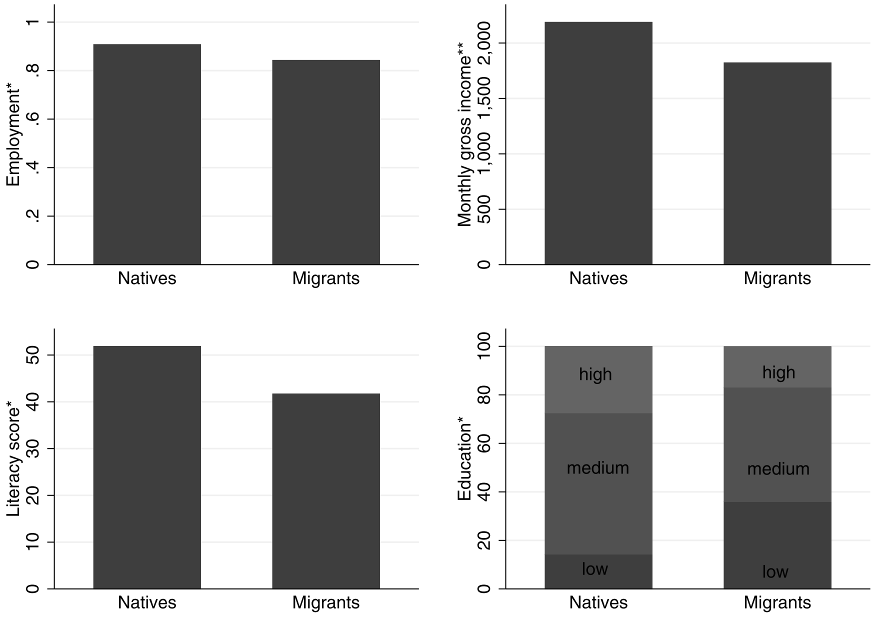

Figure 1. Comparison of migrant and native labor force

Source: LEO–Level One Study, 2010, own calculations.

Note: Averaged shares (probability weighted) based on five plausible values: *5,651 observations including 568 migrants; **4,525 observations including 427 migrants. [Colour figure can be viewed at wileyonlinelibrary.com]

income, we have imputed predicted gross wages based on linear wage regressions using respondents in the respective class who provided an exact income. The predictions are based on age, sex, education, occupation, working hours, and region.

2.2. *A Descriptive Comparison of the Migrant and Native Labor Forces*

Based on our definition of migrant status, in 2010 about 16 percent of the migrant and around 9 percent of the native labor force were unemployed (see Figure 1 and Table A.2). On average, the monthly gross wages of natives were about €366 higher than those of migrants, and we find virtually no difference in hours worked per week.

Figure A.3 shows that the literacy skills of the migrant and native workforces are widely different. It plots the density distributions of the literacy scores and shows that, compared to migrants, natives have higher reading and writing abilities. Most natives score between 40 and 65 points, whereas most migrants obtain between 30 and 50 points. The means of the two groups are 40.96 and 51.36 (see Figure 1 and Table A.2).[8]

Using discrete literacy levels, we find that strict illiteracy (functional illiteracy) affects 17.6 percent (25.9 percent) of the migrants, compared to a mere 2.4

[8]The hypothesis of the equality of means between migrants and natives can be rejected on the basis of a standard t-test (p-value = 0.000). Additionally, we conduct a Kolmogorov–Smirnov two-sample test, which clearly rejects the hypothesis of the equality of the literacy distributions between the two groups (p-value = 0.000).

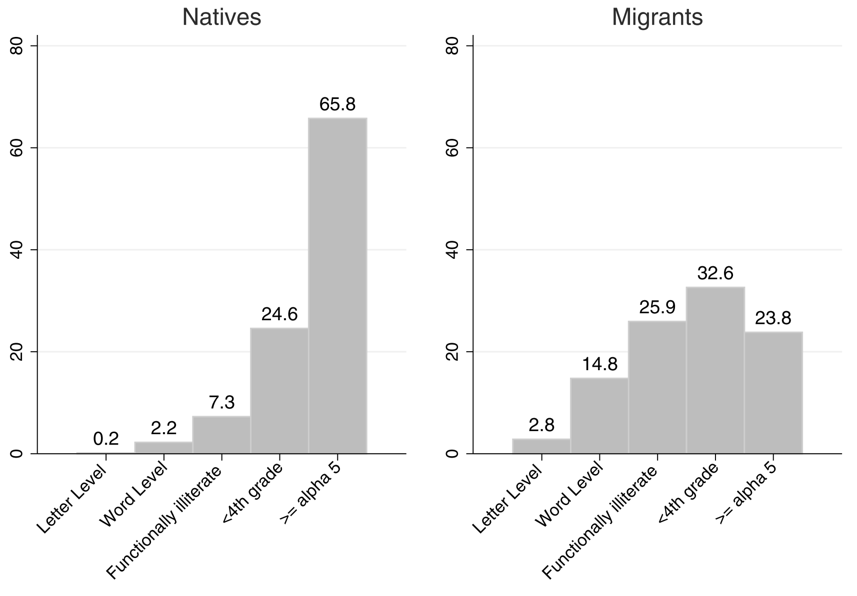

Figure 2. α-Levels in the labor force

Source: LEO–Level One Study, 2010, own calculations.
Note: Averaged shares (probability weighted) based on five plausible values: 5,651 observations including 568 migrants. [Colour figure can be viewed at wileyonlinelibrary.com]

percent (7.3 percent) among Germans (see Figure 2 and Table A.2). Furthermore, 32.6 percent (24.6 percent) of the migrants (natives) cannot read or write texts at a level that is expected at the end of fourth grade (α-level 4). Overall, the LEO survey reveals that the share of people who have adequate reading and writing skills among the migrant workforce (23.8 percent) is less than half the share among natives (65.8 percent). Moreover, the proportion of migrants who attain a test score in the range of α-level 3 or below (= cumulative functional illiteracy, 43.5 percent) is more than four times as high as the respective share among natives (9.7 percent).

As for educational differences, the prevalent view of the general public is that average school attainment among migrants is lower than among native Germans—a disparity that is often claimed to be responsible for the fact that migrants are less successful on the labor market. Figure 1 provides a glance at educational differences between migrants and natives.[9] The numbers suggest that, on average, migrants indeed possess lower educational degrees than natives: 36 percent (14.3 percent) of the migrant (native) population has at most a lower educational or occupational degree, 47.2 percent (58.2 percent) has a medium educational or occupational qualification, and 16.8 percent (27.5 percent) are highly

[9]The education variable combines the highest school and occupational degrees and generally distinguishes between three levels of education —low (≤ lower secondary education, occupational training of no more than one year), medium (= upper secondary education or post-secondary non-tertiary education, three-year occupational degree), and high (= bachelor's degree or higher, master craftsman degree).

qualified. Almost half of the immigrants (44.5 percent) have completed their highest degree abroad.

3. The Migrant–Native Literacy Gap

As stated earlier, language skills are one of the most important types of human capital because they make interacting with others possible and they facilitate the acquisition of further human capital. Natives acquire reading and writing skills (*language capital*) in their first language without much effort in the course of growing up. Migrants, on the other hand, acquire proficiency in the host-country language at a high cost.[10] These costs differ according to, for example, their educational background or the distance of the mother tongue from the destination language. While it seems obvious that, on average, migrants will have lower literacy skills in the language of the destination country, it is still informative to see to what extent the groups differ in literacy, and to investigate the determinants of literacy.

Table 1 reports results from eight weighted OLS regressions (linear probability models in columns with even numbers), where the dependent variable is either the literacy score (L. S.) or an indicator that equals one if the respondent attains a test score of α-level 3 or below ($\leq$ Funct.).[11] We estimate each specification five times—once with each relevant plausible value variable—we average the parameters, and compute clustered standard errors which are adjusted for variation between the five sets of results. All specifications include the number of years since migration (and its square), a gender dummy, a dichotomous variable for having a partner, and the number of children, as well as fixed effects for birth cohorts, population size classes, and counties (*Landkreis*). We additionally control interview duration and interviewer fixed effects to account for interviewers' potential impact on the literacy tests. We center variables which are interacted with the migrant dummy (linguistic distance and years since migration) in order to measure the literacy gap at the mean value of the interacted variables.

Column (1) of Table 1 shows that the average migrant's command of language lies about 9.6 score points ($\approx$ one standard deviation) below the linguistic abilities of an average native. In column (2), the probability of being functionally illiterate is almost 31 percentage points higher for migrants than for native speakers, for whom this probability is 9.7 percent (see Figure 2 and Table A.2). In columns (3) and (4), we additionally control for educational attainment, which reduces the literacy gap to 0.86 standard deviations of the literacy score and scales the functional illiteracy gap down to 27 percentage points (for newly arrived migrants, these gaps are of course much larger, because years since migration [and its square] is centered at its mean of roughly 22.6 years).

[10]For a comprehensive overview on the acquisition of language capital, see Chiswick (1991) and Chiswick and Miller (1995).

[11]In order to check whether the determinants of the literacy gap differ between the labor force and the entire population, we also estimate the literacy equation taking into account the full sample of the LEO study. These extended estimates very closely resemble the results for the labor force. An empirical supplement is available upon request.

TABLE 1

The Literacy Gap in the Labor Force: Language-Based Definition of Migrant. Dependent Variable: Literacy Score (L. S.)/Functional Illiteracy (≤ Funct.)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variables	L. S.	≤ Funct.	L. S.	≤ Funct.	L. S.	≤ Funct.	L. S.	≤ Funct.
Migrant	−9.558***	0.305***	−8.495***	0.273***			−8.526***	0.274***
	(0.626)	(0.036)	(0.625)	(0.036)			(0.618)	(0.036)
Afro-Asiatic languages					−9.425***	0.378***		
					(2.214)	(0.118)		
Altaic languages					−8.942***	0.312***		
					(1.133)	(0.078)		
Germanic languages					−5.550**	0.127		
					(2.391)	(0.110)		
Iranian languages					−12.785***	0.466***		
					(1.889)	(0.098)		
Romanic languages					−8.213***	0.297***		
					(1.712)	(0.107)		
Slavic languages					−7.703***	0.214***		
					(0.774)	(0.045)		
Other Indo-European languages					−9.955***	0.339***		
					(2.278)	(0.123)		
Other language groups					−10.246***	0.305**		
					(2.165)	(0.139)		
Centered linguistic distance × migrant							−0.089	0.005*
							(0.073)	(0.003)
Education medium			4.284***	−0.156***	4.183***	−0.149***	4.278***	−0.155***
			(0.404)	(0.022)	(0.401)	(0.021)	(0.403)	(0.022)
Education high			7.891***	−0.218***	7.724***	−0.209***	7.844***	−0.215***
			(0.493)	(0.024)	(0.493)	(0.023)	(0.497)	(0.024)

Table 1 *Continued*

Variables	(1) L. S.	(2) ≤ Funct.	(3) L. S.	(4) ≤ Funct.	(5) L. S.	(6) ≤ Funct.	(7) L. S.	(8) ≤ Funct.
Centered years since migration × migrant	0.118***	−0.004**	0.122***	−0.004**	0.117***	−0.005*	0.122***	−0.004**
	(0.041)	(0.002)	(0.040)	(0.002)	(0.042)	(0.002)	(0.040)	(0.002)
Male	−2.879***	0.065***	−3.186***	0.073***	−3.151***	0.069***	−3.179***	0.072***
	(0.368)	(0.012)	(0.371)	(0.012)	(0.375)	(0.012)	(0.371)	(0.012)
Has partner	1.121**	−0.034**	0.844**	−0.027*	0.806*	−0.026*	0.834**	−0.027*
	(0.444)	(0.017)	(0.409)	(0.016)	(0.412)	(0.016)	(0.411)	(0.016)
Number of children ≤ 6 years	−0.051	0.008	0.050	0.004	0.093	0.001	0.061	0.003
	(0.429)	(0.013)	(0.419)	(0.013)	(0.414)	(0.013)	(0.419)	(0.013)
Number of children. 7–13 years	0.104	0.010	0.256	0.006	0.283	0.004	0.266	0.005
	(0.276)	(0.014)	(0.267)	(0.014)	(0.264)	(0.013)	(0.266)	(0.014)
Number of children 14–17 years	−0.473	0.013	−0.344	0.010	−0.323	0.008	−0.323	0.008
	(0.475)	(0.015)	(0.451)	(0.014)	(0.445)	(0.015)	(0.446)	(0.014)
Constant	55.248***	−0.061	49.398***	0.126	49.687***	0.119	49.551***	0.118
	(4.064)	(0.125)	(3.960)	(0.126)	(4.082)	(0.127)	(3.941)	(0.125)
Cohort fixed effects χ^2_{46}	59.16*	80.17***	91.25***	92.18***	89.81***	93.14***	91.52***	93.16***
Population size fixed effects χ^2_{6}	1.022	0.343	1.276	0.425	1.489	0.409	1.367	0.476
County fixed effects χ^2_{83}	2,031***	1,170***	1,916***	1,199***	1,845***	1,045***	1,902***	1,129 ***
Interviewer fixed effects χ^2_{281}	1.647×10^6***	6.147×10^6***	1.075×10^6***	4.613×10^6***	3.644×10^6***	5.217×10^6***	1.676×10^6***	3.490×10^6***

Source: LEO–Level One Study, 2010, own calculations.

Notes: Averaged parameters from five weighted OLS estimates based on the relevant plausible value: 5,651 observations including 568 migrants. Dependent variable: (non-standardized) literacy score in odd-numbered columns, an indicator for being functionally illiterate (α-level 3 or below) in the even-numbered columns. All regressions additionally controled for interview duration and for cohort, population size, county, and interviewer fixed effects. Standard errors in parentheses clustered by county and adjusted for variation between the five sets of coefficients: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

The degree of difficulty in acquiring the host-country language varies depending on the migrant's first language (see, e.g., Chiswick and Miller, 2005, 2012, 2015; Isphording and Otten, 2014). We capture this in two ways: In columns (5) and (6), we include a set of self-constructed binary variables which classify the immigrants' native tongues according to "language family trees." Since Chiswick and Miller (2005) argue that the use of language trees may not fully cover how a modern language differs from (1) its predecessor language, (2) other language branches on the same tree, and (3) modern languages on other trees, we additionally use a recently developed continuous linguistic distance measure in columns (7) and (8). The distance measure is provided by the German Max Planck Institute for Evolutionary Anthropology, and uses an algorithm to measure similarity in pronunciation and vocabulary of languages—in our case, the similarity of German with other languages (see Bakker *et al.*, 2009; Wichmann *et al.*, 2016). Greater distances to a language are thus associated with greater difficulty and a higher cost of learning that language (for more details, see Isphording and Otten, 2014; Chiswick and Miller, 2015). *Ceteris paribus*, the literacy gap is smallest for migrants with Germanic (0.56 standard deviations and 13 percentage points), Slavic (0.78 standard deviations and 21 percentage points), and Romanic (or "Romance") (0.83 standard deviations and 30 percentage points) language backgrounds. The largest difference is found with respect to the Iranian language tree (1.3 standard deviations and 47 percentage points). As for the distance measure, we find that an increase of linguistic distance by one standard deviation raises the language gap by 0.06 standard deviations of the literacy score and increases the probability gap of being functionally illiterate by 3.5 percentage points. These results show that while there is a large gap in literacy between the general group of migrants and natives, the variation of literacy within the migrant group is not as large.

Coefficients for most of the other variables are as expected. Exposure to the host-country language approximated by the number of years since migration is positively correlated with higher language skills—an additional year in the host country decreases the literacy score gap (probability gap of being functionally illiterate) by about 0.01 standard deviations (between 0.4 and 0.5 percentage points). Being male is associated with poorer reading and writing abilities, and having a partner is correlated with better literacy, whereas the number of children does not seem to be linked to the ability to read and write.

4. The Migrant–Native Employment Gap

Having established that there are substantial differences in literacy between migrants and natives, we investigate whether the discrepancies are related to labor market outcomes. The ability to read and write fluently is not only expected to affect wages—more generally, it is usually also a prerequisite for employment. Literacy may, for example, be a decisive factor in finding out about vacancies or convincing potential employers in job interviews. Furthermore, in many work environments it is only through literacy that other forms of human capital become usable. Even more so than for natives, this argument applies to migrants.

TABLE 2

THE EMPLOYMENT GAP IN THE LABOR FORCE: LANGUAGE-BASED DEFINITION OF MIGRANT. DEPENDENT VARIABLE: PART-/FULL-TIME EMPLOYED VS. UNEMPLOYED

Variables	(1)	(2)	(3)	(4)	(5)	(6)
Migrant	−0.056***	−0.032**	0.000	0.003	0.002	0.008
	(0.014)	(0.015)	(0.016)	(0.017)	(0.017)	(0.019)
Standardized literacy score			0.039***		0.039***	
			(0.006)		(0.006)	
Alpha 1 or 2				−0.146***		−0.161***
				(0.042)		(0.055)
Functionally illiterate				−0.101***		−0.112***
				(0.026)		(0.027)
<Fourth-grade level literacy				−0.037***		−0.036***
				(0.012)		(0.013)
Standardized literacy score × migrant					0.002	
					(0.017)	
Centered Alpha1/2 × migrant						0.043
						(0.072)
Centered functionally illiterate × migrant						0.046
						(0.053)
Centered <fourth grade × migrant						0.008
						(0.041)
Education medium		0.126***	0.109***	0.106***	0.109***	0.106***
		(0.017)	(0.017)	(0.016)	(0.017)	(0.017)
Education high		0.159***	0.125***	0.127***	0.125***	0.127***
		(0.019)	(0.018)	(0.018)	(0.018)	(0.018)
Centered years since migration × migrant	0.003**	0.003**	0.003**	0.003**	0.003**	0.003**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Centered linguistic distance × migrant			−0.004**	−0.004**	−0.004**	−0.004**
			(0.002)	(0.002)	(0.002)	(0.002)
Male	−0.055***	−0.057***	−0.045**	−0.047**	−0.045**	−0.046**
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
Has partner	0.087***	0.084***	0.079***	0.079***	0.079***	0.079***
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)

Table 2 *Continued*

Variables	(1)	(2)	(3)	(4)	(5)	(6)
Partner × male	0.051**	0.047**	0.048**	0.048**	0.048**	0.047**
	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)
Number of children ≤ 6 years	−0.033***	−0.029***	−0.028***	−0.028***	−0.028***	−0.028***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Number of children 7–13 years	−0.009	−0.006	−0.007	−0.006	−0.007	−0.006
	(0.009)	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)
Number of children 14–17 years	−0.016	−0.014	−0.012	−0.012	−0.012	−0.012
	(0.012)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Constant	0.930***	0.783***	0.790***	0.815***	0.790***	0.815***
	(0.063)	(0.066)	(0.066)	(0.068)	(0.066)	(0.068)
Cohort fixed effects χ^2_{46}	108.5***	130.6***	94.48***	94.23***	93.70***	94.17***
Population size fixed effects χ^2_{6}	18.38***	18.90***	19.09***	18.48	19.09***	18.36***
County fixed effects χ^2_{83}	2.1×10^7***	3.03×10^7 ***	26,127.7***	22,936.1***	23,714.08***	20,457.23***
Interviewer fixed effects χ^2_{281}	8.4×10^5***	3.12×10^6***	$1.59{\times}10^8$***	$2.03{\times}10^8$***	$1.24{\times}10^8$***	5.51×10^7***

Source: LEO–Level One Study, 2010, own calculations.

Notes: Averaged parameters from five weighted OLS estimates based on the relevant plausible value: 5,651 observations including 568 migrants and 4,525 employed respondents. All regressions additionally controled for interview duration and for cohort, population size, county, and interviewer fixed effects. Standard errors in parentheses clustered by county and adjusted for variation between the five estimates: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

In their case, even those with high levels of human capital acquired in the country of origin may find it difficult to find an appropriate job in the host country. In addition, for low-skilled jobs also it is important to have a sufficient command of the host-country language in order, for example, to follow work instructions or to comply with the employer's health and safety regulations.

Table 2 sheds light on the employment gap between migrants and natives and its link to language proficiency. All specifications include the following covariates: centered number of years since migration, a gender dummy, a dichotomous variable for having a partner, the number of children, and cohort, population size, and county fixed effects. We control for the effect that interviewers may have on the literacy tests by including interview duration and interviewer dummies. We only consider people who are in the labor force and we set the dependent variable equal to one for those who are employed (full- or part-time) and to zero for those who are currently unemployed. The number of years since migration and linguistic distance are centered and set to zero for natives, that is, interacted with the migrant dummy. To fully exploit statistical efficiency gains from the plausible values, we estimate each specification five times using computationally undemanding and readily comparable linear probability models.

We conduct a stepwise approach of adding further control variables through columns (1)–(6). Column (1) shows that—without educational and linguistic controls—the proportion of natives who work is 5.6 percentage points larger than the proportion of migrants (mean employment share among natives: 90.8 percent). Correcting for the fact that, on average, migrants and natives possess different educational degrees in column (2) reduces the difference to 3.2 percentage points. Finally, controlling for the ability to read and write in columns (3)–(6) virtually eliminates the migrant–native employment gap—that is, the coefficient of the migrant dummy is close to zero and no longer significant, and thus the employment gap can be fully explained by the literacy gap. When interpreting the coefficients of the controls, one should keep in mind that those with better reading and writing abilities are also more likely to search for a job, and thus self-selection may explain the results to some extent. This is true for both migrants and natives, and so should not matter much for our assessment of the employment gap.

In columns (3) and (4), we restrict the relationship between literacy and employment to be the same for migrants and natives. An increase in the literacy score by one standard deviation increases the probability of being employed by 3.9 percentage points (column 3).[12] As for the discrete literacy levels in column (4), the employment probability of literates differs from individuals at α-level 1 or 2 by 14.6 percentage points, and narrows to 10.1 (3.7) percentage points for individuals at α-level 3 (α-level 4).

In columns (5) and (6), we apply a more flexible approach and allow the literacy parameters to vary between natives and migrants. However, the interaction terms between the migrant indicator and the literacy variables are insignificant in both specifications, and the parameters for natives are almost the same as in the restricted estimates. This is important because it suggests that literacy does indeed

[12]We use standardized literacy scores in all employment and wage regressions; literacy score coefficients can therefore readily be interpreted in terms of standard deviations.

measure the same thing for both migrants and natives. The literacy coefficient may still capture elements of ability for both groups alike (thus precluding us from causally estimating the returns to literacy). What matters for our statements about the wage gap, however, is that the coefficient is not *differentially* confounded with motivation or ability for the two groups.

The literacy score coefficient may also differ from the true causal effect of literacy if being employed affects language proficiency. Such reverse causality again prevents a clean identification of the returns to literacy, but for our purposes reverse causality is only an issue if it differentially affects migrants and natives; for example, if employment increases literacy skills especially for migrants. Reverse causality could then in part be responsible for the reduction of the employment gap that we observe when controlling for literacy. If the true returns to literacy are the same for migrants and natives, we should in that case find that the coefficients for the literacy score differ between migrants and natives. As we have already shown, the interaction term is not significant and therefore provides no indication that reverse causality drives our results concerning the employment gap.

We further control for the linguistic distance between mother tongue and German in specifications (3)–(6), and find that a greater distance significantly reduces the probability that a migrant is in employment. Since we are already holding literacy constant, the distance coefficient cannot be driven by different language skills. Rather, it seems plausible that linguistic distance may in that case also capture characteristics such as differences in work ethics, missing networks, cultural differences, or informational deficiencies with respect to the host-country labor market.

The coefficients of the years since migration are the same in all specifications. An additional year in the host country increases the probability of being employed by 0.3 percentage points. Interestingly, the coefficient on time since migration does not seem to be driven by migrants' improvements in literacy, as it remains unchanged when adding literacy as controls. Restricting the coefficients to be the same for migrants and natives, we find that women in the labor force have a 5 percentage points higher probability of being employed. Having a partner increases the probability of working by 9 percentage points for females and about 14 percentage points for males. The number of small children (less than seven years old), on the other hand, is negatively correlated with the probability of working, whereas older children (aged between seven and 17) do not seem to make a difference. The education parameters are significantly positive and, as expected, the magnitude increases with the education level. Overall, our results for the control variables support the results of other studies in, for example, the U.K. (Dustmann and Fabbri, 2003) and Germany (Jäckle and Himmler, 2010).

5. The Migrant–Native Wage Gap

The monthly wages of migrants in our sample are on average €366 lower than those of natives (see Section 2.2). There are a variety of reasons for the wage gap. For example, lower earnings may be explained by lower educational degrees, missing networks, and informational deficiencies with respect to the host-country labor market, but of course also by poor command of the host-country language. Furthermore, because literacy skills are complementary to any human capital acquired in the country of origin, this human capital is usually not perfectly transferable to the host country. As migrants make investments to learn the foreign language and to improve the transferability of their human capital, the costs of these investments may temporarily actually have a negative effect on earnings and slow any wage assimilation. In the course of time the wage gap should, however, become smaller because the extent of investments in language acquisition decreases, and the earlier investments in language skills pay off by allowing individuals to better utilize their human capital on the labor market.

Table 3 presents six Mincer (1958, 1974) wage regressions, where the dependent variable is the log of gross monthly wages. We control the log of working hours per week, and two dummy variables indicating whether the individual is working part-time or is self-employed. In order to account for the fact that in the course of time migrants adapt to the host country in respects other than language, we include the number of years since migration; finally, we also add linguistic distance. The latter two variables are centered at their mean and set to zero for natives. Making use of the plausible values in the LEO dataset, we estimate each specification five times to reduce the measurement error in the command of language of migrants and natives to a minimum.

All standard controls have the expected signs and are estimated to be statistically significant at the 1 percent level: on average, individuals who are better educated, employees who work longer hours, men, and individuals in a partnership have higher salaries, while part-time employees and those who are self-employed earn lower monthly wages. Also, both the number of years since migration and the linguistic distance are positively correlated with wages, but are not statistically significant. As linguistic distance may already capture network effects and differences in cultural dimensions, as well as informational deficiencies, we are confident that the parameters of the literacy variables are not confounded with these factors. Restricting the relationship between literacy and wages to be the same for migrants and natives, we find that an increase in the literacy score by one standard deviation increases wages by 7.2 percent (column 3). The results in column (4) show that wages of literates differ from those of individuals on α-level 1 or 2, α-level 3, and α-level 4 by 27 percent, 19.2 percent, and 7.4 percent. Allowing the literacy coefficients to vary between migrants and natives in columns (5) and (6) does not change any of the results and all interactions are insignificant, suggesting that the restricted specifications are already valid.

In the first column of Table 3, we do not include linguistic or educational control variables and we find that migrants earn on average 14.6 percent less than natives. Based on an average salary for natives of €2,189, the wage gap is €320. Column (2) demonstrates that about half of the earnings differential reflects differing educational levels of migrants and natives. Conditional on the literacy variables in columns (3)–(6), however, the wage gap vanishes. This result is similar to what we find in Table 2 when analyzing employment probabilities. Both the wage gap and employment differences between migrants and natives can be fully explained by the literacy gap.

TABLE 3

The Wage Gap Amongst Employees: Language-Based Definition of Migrant. Dependent Variable: Log Wages

Variables	(1)	(2)	(3)	(4)	(5)	(6)
Migrant	−0.146***	−0.076***	−0.014	−0.010	0.000	−0.008
	(0.029)	(0.028)	(0.030)	(0.029)	(0.031)	(0.032)
Standardized literacy score			0.072***		0.069***	
			(0.016)		(0.016)	
Alpha 1 or 2				−0.270***		−0.304***
				(0.084)		(0.114)
Functionally illiterate				−0.192***		−0.182***
				(0.054)		(0.053)
<fourth-grade level literacy				−0.074***		−0.078***
				(0.024)		(0.027)
Standardized literacy score × migrant					0.025	
					(0.032)	
Centered Alpha1/2 × migrant						0.073
						(0.144)
Centered functionally illiterate × migrant						−0.017
						(0.103)
Centered <fourth grade × migrant						0.030
						(0.081)
Education medium		0.251***	0.223***	0.217***	0.221***	0.217***
		(0.028)	(0.026)	(0.027)	(0.026)	(0.027)
Education high		0.568***	0.518***	0.520***	0.516***	0.520***
		(0.035)	(0.034)	(0.034)	(0.034)	(0.033)
Centered time since migration × migrant	0.004	0.004	0.003	0.003	0.003	0.003
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Centered linguistic distance × migrant			0.005	0.006	0.006	0.006
			(0.004)	(0.004)	(0.004)	(0.004)
Log working hours	0.875***	0.843***	0.844***	0.844***	0.844***	0.843***
	(0.054)	(0.047)	(0.047)	(0.047)	(0.047)	(0.046)
Male	0.191***	0.190***	0.213***	0.210***	0.213***	0.210***
	(0.021)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)

Table 3 *Continued*

Variables	(1)	(2)	(3)	(4)	(5)	(6)
Has partner	0.082***	0.070***	0.068***	0.069***	0.069***	0.068***
	(0.020)	(0.020)	(0.021)	(0.021)	(0.021)	(0.021)
Part-time employed	−0.281***	−0.256***	−0.254***	−0.255***	−0.254***	−0.255***
	(0.047)	(0.044)	(0.044)	(0.044)	(0.043)	(0.043)
Self-employed	−0.144***	−0.161***	−0.158***	−0.160***	−0.158***	−0.161***
	(0.036)	(0.038)	(0.038)	(0.038)	(0.038)	(0.038)
Constant	4.425***	4.128***	4.102***	4.152***	4.103***	4.154***
	(0.218)	(0.192)	(0.190)	(0.193)	(0.190)	(0.192)
Cohort fixed effects χ^2_{46}	1,065***	729.1***	724.0***	680.2***	695.6***	629.9***
Population size fixed effects χ^2_{6}	6.664	5.677	5.567	5.594	5.624	5.560
County fixed effects χ^2_{80}	1.29×10^6***	1.79×10^6***	11,767***	9,478***	10,257***	10,497***
Interviewer fixed effects χ^2_{265}	2.95×10^6***	10,3251***	3.7×10^7***	3.01×10^7***	2×10^7***	5.16×10^7***

Source: LEO–Level One Study, 2010, own calculations.

Notes: Averaged parameters from five weighted OLS estimates based on the relevant plausible value: 4,525 observations including 427 migrants. All regressions additionally controled for interview duration and for cohort, population size, county, and interviewer fixed effects. Standard errors in parentheses clustered by county and adjusted for variation between the five estimates: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Just as in the context of the employment equation, the interaction term between literacy score and migrant status is not statistically significant. In that respect, our estimates provide no indication that unobserved ability and reverse causality are responsible for the reduction of the wag gap that we observe when controlling for literacy. In addition, Figure A.1 compares our literacy coefficient with the returns to literacy that have been reported in other studies. In the top panel, we display our estimate from column (3), which suggests that a one standard deviation increase in literacy is associated with 7.2 percent higher wages. This is very close to the number that Dustmann and van Soest (2001) report in their instrumental variable estimations as the causal returns to literacy in Germany (7.3 percent). Further, we compare our results with those of Bleakley and Chin (2004), who in their seminal paper identify the causal effect of literacy from variation in age at arrival of migrants hailing from English-speaking and non-English-speaking countries. Their estimations do not control for education and suggest that a one standard deviation increase in literacy causally increases wages by 17 percent in the U.S. When we also omit the education controls for the sake of comparison, the estimate we obtain is much lower, at 11.7 percent. The figures also show our estimates of the wage gap from simulations which fix the returns to literacy at different levels. As can be seen, the wage gap would be similar or even smaller in our data at the returns to literacy reported by Dustmann and van Soest (2001) and Bleakley and Chin (2004).

The magnitudes of our coefficient estimates therefore do not particularly suggest that they are driven upwards by reverse causality or unobserved heterogeneity. This is consistent with the literature, which finds that OLS estimates are typically (much) lower than instrumental variable estimates, and interprets this as evidence that downward bias from measurement error affects literacy coefficients much more than upward bias due to reverse causality or unobserved heterogeneity (see also Dustmann and Glitz, 2011). Together with the finding that there is no interaction effect between being a migrant and the literacy score, this gives us confidence that most of the wage gap can, in fact, be explained by factors relating to productivity.

Table A.5 presents some consistency checks considering different specifications of the wage equation. First, in order to account for different professions, we construct four occupation dummies: high-skilled white-collar, low-skilled white-collar, high-skilled blue-collar, and low-skilled blue-collar.[13] The estimated parameters of these dummy variables in columns (4)–(6) have the expected signs. They are statistically significant at the 1 percent-level. On average, high-skilled white-collar workers have the highest earnings, and low-skilled white-collar and high-skilled blue-collar employees command similar wages, whereas low-skilled blue-collar employees earn the least. Compared to our baseline results in columns (1)–(3), we find that the parameter of the literacy score (column 6) and the wage gaps in columns (4)–(6) are smaller. A possible explanation for this could be that

[13]The classifications are based on the definitions of the European Foundation for the Improvement of Living and Working Conditions (Eurofound); see http://www.eurofound.europa.eu/surveys/ewcs/2005/classification. For a detailed description of the categories, see Table A.1; summary statistics are presented in Table A.2.

individuals select different occupations based (in part) on their linguistic abilities. That is to say, the occupation dummies are not independent of literacy and, therefore, capture part of the literacy effect when added as controls. This is the case for both natives and migrants, but in the case of migrants it may be even more pronounced, as being able to read and write in the host-country language is complementary to any professional skills that immigrants may have acquired in their home countries. Additionally, in order to allow the literacy variable to vary across different occupations, we extend the specification in column (6) using interaction terms. The interactions are insignificant for all occupation categories, which suggests that literacy is associated with similar wage premiums within the professional categories.

In columns (7)–(12) of Table A.5, we use more flexible functional forms of the variable "time since migration". In columns (10)–(12), we include a full set of dummy variables for each year of residence. Furthermore, we test a non-linear specification with "years since migration squared" as an additional control variable in columns (7)–(9). The coefficient of the squared term is negative but statistically insignificant. Using the more flexible fixed-effects approach, we find that the dummy variables are jointly significant. Most importantly, we do not find any substantial changes in the results when we compare our initial linear "time since migration" specification to these alternative functional forms.

6. Robustness and Extensions

This section explores how robust our results are with respect to the use of alternative definitions of migrant status, and across subgroups of migrants. In addition, we investigate whether our estimates are heterogeneous across subsamples defined by individual characteristics such as gender or age at migration. The results of these extensions and consistency checks are displayed in Tables 4 and 5. For reference, the first three columns of both tables repeat the baseline estimates from the wage equation in the previous section (Table 3).

6.1. *Alternative Definitions of Migrant Status*

So far, we have used a definition of migrant status that is based on the mother tongue of the respondent. We check the robustness of our results in Table 4 by using two other, more commonly used, definitions. The first definition is being "born abroad" (columns 4–6). In comparison with our language-based definition, this reassigns those born in Germany who do not report German to be their mother tongue from the migrant to the native group. At the same time, it reassigns those born abroad who report German to be their mother tongue from the native to the migrant group. The second definition is "foreign citizenship" (columns 7–9). In comparison with our language-based definition, this reassigns German citizens who do not report German to be their mother tongue from the migrant to the native group. At the same time, it assigns foreign citizens who report German to be their mother tongue from the native to the migrant group. In weighted numbers, 9.4 percent of the labor force and 8.7 percent of the employed are foreign citizens, and 15.8 percent of the labor force and 14.9 percent

TABLE 4

The Wage Equation: Extensions and Robustness Checks (I). Dependent Variable: Log Wages

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Full Sample: Language-Based Definition			Foreign-Born			Foreign Citizens			Born Abroad—GER Mother Tongue		
Migrant	−0.146***	−0.076***	−0.014	−0.168***	−0.097***	−0.041	−0.107**	−0.032	0.018	−0.108*	−0.053	−0.018
	(0.029)	(0.028)	(0.030)	(0.035)	(0.033)	(0.034)	(0.048)	(0.044)	(0.042)	(0.057)	(0.051)	(0.050)
Standardized literacy score			0.072***			0.068***			0.073***			0.065***
			(0.016)			(0.016)			(0.016)			(0.016)
Education medium		0.251***	0.223***		0.248***	0.221***		0.257***	0.226***		0.271***	0.246***
		(0.028)	(0.026)		(0.028)	(0.027)		(0.028)	(0.027)		(0.035)	(0.034)
Education high		0.568***	0.518***		0.564***	0.514***		0.576***	0.518***		0.606***	0.559***
		(0.035)	(0.034)		(0.034)	(0.034)		(0.035)	(0.035)		(0.039)	(0.039)
Observations	4,525	4,525	4,525	4,525	4,525	4,525	4,525	4,525	4,525	4,098	4,098	4,098
Observed migrants	427	427	427	522	522	522	201	201	201	149	149	149

Variables	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)
	Small Linguistic Distance			Large Linguistic Distance			Migrant Networks			EU Citizens		
Migrant	−0.111***	−0.065*	−0.016	−0.189***	−0.079*	−0.012	−0.143***	−0.046	0.004	−0.109	−0.046	−0.025
	(0.039)	(0.037)	(0.039)	(0.051)	(0.047)	(0.046)	(0.038)	(0.035)	(0.037)	(0.098)	(0.101)	(0.096)
Standardized literacy score			0.065***			0.072***			0.066***			0.067***
			(0.016)			(0.017)			(0.015)			(0.016)
Education medium		0.258***	0.232***		0.260***	0.232***		0.267***	0.239***		0.258***	0.222***
		(0.029)	(0.028)		(0.032)	(0.031)		(0.032)	(0.031)		(0.028)	(0.027)

Table 4 *Continued*

	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)
Variables	Small Linguistic Distance			Large Linguistic Distance			Migrant Networks			EU Citizens		
Education high		0.588***	0.540***		0.582***	0.530***		0.602***	0.552***		0.583***	0.526***
		(0.035)	(0.034)		(0.037)	(0.036)		(0.037)	(0.036)		(0.033)	(0.033)
Observations	4,356	4,356	4,356	4,267	4,267	4,267	4,286	4,286	4,286	4,413	4,413	4,413
Observed migrants	258	258	258	169	169	169	188	188	188	89	89	89

Source: LEO–Level One Study, 2010, own calculations.

Notes: Averaged parameters from five weighted OLS estimates, each one calculated using the relevant plausible value. All specifications include log hours of work, interview duration, a dummy for being male, having a partner, being part-time or self-employed, and a full set of cohort, population size, county, and interviewer fixed effects. Columns (1)–(3) repeat the baseline results of Table 3. The estimates in columns (4)–(9) use alternative definitions of migrant. The results in columns (10)–(12) are based on a subsample of the foreign-born migrant definition in columns (4)–(6). Columns (13)–(21) use the language-based definition in columns (1)–(3) and the results in columns (22)–(24) are based on the foreign citizen sample. Standard errors in parentheses clustered by county and adjusted for variation between the five estimates: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE 5

The Wage Equation: Extensions and Robustness Checks (II). Dependent Variable: Log Wages

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Variables	Full Sample: Language-Based Definition			Full- Time			Male			Female		
Migrant	−0.146***	−0.076***	−0.014	−0.183***	−0.100**	−0.046	−0.228***	−0.130***	−0.067	−0.075	−0.038	0.023
	(0.029)	(0.028)	(0.030)	(0.048)	(0.047)	(0.045)	(0.041)	(0.039)	(0.041)	(0.057)	(0.052)	(0.056)
Standardized literacy score			0.072***			0.063***			0.074***			0.066**
			(0.016)			(0.017)			(0.024)			(0.028)
Education medium		0.251***	0.223***		0.246***	0.224***		0.255***	0.223***		0.245***	0.222***
		(0.028)	(0.026)		(0.037)	(0.035)		(0.056)	(0.052)		(0.044)	(0.043)
Education high		0.568***	0.518***		0.546***	0.500***		0.555***	0.498***		0.590***	0.550***
		(0.035)	(0.034)		(0.047)	(0.043)		(0.073)	(0.067)		(0.053)	(0.055)
Observations	4,525	4,525	4,525	3,107	3,107	3,107	2,200	2,200	2,200	2,325	2,325	2,325
Observed migrants	427	427	427	291	291	291	237	237	237	190	190	190

	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)
Variables	West Germany			Age at Migration >11			Age at Migration ≤11		
Migrant	−0.158***	−0.090***	−0.034	−0.184***	−0.099***	−0.033	−0.061	−0.013	0.023
	(0.027)	(0.027)	(0.028)	(0.040)	(0.036)	(0.039)	(0.050)	(0.063)	(0.059)
Standardized literacy score			0.064***			0.067***			0.071***
			(0.016)			(0.017)			(0.016)

Table 5 *Continued*

	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)
Variables	West Germany			Age at Migration >11			Age at Migration ≤11		
Education medium		0.252***	0.227***		0.267***	0.240***		0.243***	0.217***
		(0.030)	(0.028)		(0.030)	(0.029)		(0.033)	(0.032)
Education high		0.567***	0.523***		0.594***	0.545***		0.569***	0.520***
		(0.036)	(0.036)		(0.036)	(0.035)		(0.038)	(0.038)
Observations	3,488	3,488	3,488	4,393	4,393	4,393	4,230	4,230	4,230
Observed migrants	395	395	395	295	295	295	132	132	132

Source: LEO–Level One Study, 2010, own calculations.

Notes: Averaged parameters from five weighted OLS estimates, each one calculated using the relevant plausible value. All specifications include log hours of work, interview duration, a dummy for being male, having a partner, being part-time or self-employed, and a full set of cohort, population size, county, and interviewer fixed effects. Columns (1)–(3) repeat the baseline results of Table 3. Estimates in columns (4)–(21) use subamples of the language-based migrant definition in columns (1)–(3). Standard errors in parentheses clustered by county and adjusted for variation between the five estimates: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

of the employed were not born in Germany (see the summary statistics in Table A.4). The corresponding numbers for our "mother tongue" definition—which are not provided by the Statistical Office—are 13.6 percent and 12.8 percent.

Compared to the language-based definition, the raw wage gap for those born abroad has roughly the same size, at 16.8 percent (see Table 4, column 4). This is not so surprising, when taking into account that of the 427 migrants in the language-based definition, 373 are also classified as migrants under the "born abroad" definition, that is, there is a large overlap (Table A.3 displays the corresponding cross-tabulations). Column (5) shows that about half of the gap is due to educational differences, and the gap vanishes fully when we additionally control for literacy (column 6).

Using foreign citizenship as the criterion for migrant status in columns (7)–(9), we find a smaller raw wage gap than in the estimations based on our language definition. This is a consequence of less overlap between the definitions: only 164 of the 427 language definition migrants are also classified as migrants under the citizenship definition (see the cross-tabulations in Table A.4). Those German citizens who do not report German as their mother tongue have low literacy scores (on average 42.5, only slightly higher than the average of all non-native German speakers) and wages (on average 7.33 log points, which is roughly the average of all non-native German speakers). These individuals are now reassigned to the native group, thus lowering the average literacy score and wages there. At the same time, those with a German mother tongue but no German citizenship have higher literacy scores (on average, 47.7) and higher wages (on average, 7.41 log points) than the average migrant in our initial language-based definition—they are now reassigned from the native to the migrant group. The combined result is a smaller raw wage gap. However, it is important to stress that the patterns are similar across the definitions: Here too, when controlling for educational attainment, the gap decreases by more than half and it disappears completely with literacy as an additional control. Overall, this confirms the robustness of our initially reported results to the use of alternative definitions of migrant status.

6.2. *Subsamples of Migrants*

"German-Speaking Migrants"

Another robustness check that we implement is the estimation of wage gaps defining as migrants only individuals who were born abroad but who still report that they grew up with German as (one of) their mother tongue(s). The idea is that this provides an alternative approach to netting out the effects of language proficiency, which can be implemented even in the absence of test score data.[14] Column (10) in Table 4 shows that this results in a smaller, but still significant raw wage gap.

[14]We compare this new migrant sample to individuals who were born in Germany and report that they grew up with German as their first language, or bilingually with German as one of the languages. We exclude from both the native and migrant group those individuals who do not report German as (one of) the language(s) with which they grew up. Note that this approach estimates the wage gap between two groups of individuals, all of whom would be in the native group under our language-based migrant definition.

Our data allow us to investigate whether this remaining raw wage gap is really due to factors other than language and education. We find that the language skills of the "German-speaking migrants" are still lower than the linguistic abilities of German speakers who were born in Germany. The difference in literacy scores is 6.1 points (51.7 vs. 45.6 points; approximately two thirds of a standard deviation). In comparison, the difference between our initial migrant definition—those who do not report their native tongue to be German—and native German speakers is 10.1 score points (see Table A.2). While the raw wage gap does not disappear, we still observe the usual pattern in columns (11) and (12). About half of the earnings gap emerges due to differing education levels. Adding the literacy variable, the wage differential disappears almost completely.

The persisting raw wage gap in combination with the remaining differences in literacy suggest that it can be misleading to assume that foreign-born individuals who self-report to speak the host-country language are equally as proficient in that language as native-born speakers. The results therefore underscore the advantages of test scores over the use of self-reported indicators of language proficiency.

Linguistic Distance, Ethnic Networks, and EU Citizens

In columns (13)–(18), we test whether our results differ across levels of linguistic distance. In columns (13)–(15), we include only migrants whose native language is close to German. Columns (16)–(18) show estimates for distant languages.[15] The results are as expected: for migrants whose linguistic distance to German is small, the raw wage gap (without any educational or linguistic controls) is smaller than in the group of migrants with distant languages. However, this effect is less pronounced when we control for education, and it disappears when we add the literacy score as an additional control variable. Despite the differences in raw wage gaps, both can be explained by the education and literacy levels of the individuals in the two groups.

Having access to a network of individuals that speak a migrant's native language may also affect the extent to which proficiency in the host-country language matters. Our data lack information on ethnic networks, but we can restrain the migrant sample to the two languages that are most prevalent: Turkish and Russian (columns 19–21). The assumption here is that people who report their native tongue to be Russian or Turkish also possess the best networks, due to the large communities in Germany that speak these languages. Our (weighted) data show that 41.8 percent of the migrants in the labor force name Russian or Turkish as their mother tongue (employed migrants: 42.0 percent). We exclude migrants from the sample who name any other non-German language as their mother tongue. The raw wage gap is very similar to the gap based on our initial sample.

[15]We base the threshold for the distinction between close and distant languages on the results in Table 1, which suggest that the literacy gap is smallest for migrants from the Germanic, Slavic, and Romanic "language family trees." Accordingly, we set the linguistic distance threshold at 97.01, which is the highest value of the Romanic languages. Of the migrants in the labor force, 45.9 percent belong to the Germanic, Slavic, or Romanic language groups; among the employed migrants, 54.5 percent belong to these groups (weighted numbers).

However, if we additionally control for educational attainment, the migrant coefficient becomes insignificant and is a little smaller than in the full sample. The wage gap disappears completely if we include the literacy score. Overall, these results do not provide compelling evidence that speaking the native language of the two largest ethnic minorities in Germany is associated with a smaller wage gap. It is likely that networks still matter, but our measure is too crude to actually detect these effects.

Finally, we consider (pre-expansion, 2013) EU citizens as a set of migrants who can potentially find a job more easily (columns 22–24). As our sample includes only individuals who are either employed or unemployed, all migrants in this sample do in fact possess a work permit. From that perspective, the EU citizenship should not make it easier to find employment. However, there may be other factors, such as cultural differences or informational deficiencies, which attach migrants from the EU more closely to the labor market and allow them to earn higher wages than other migrants—even if all migrants are permitted to work. The raw earnings gap for EU citizens is 10.9 percent, which is hardly different from the sample of all foreign citizens (columns 7–9). Again, about half of the gap arises due to educational differences, and the gap vanishes fully if we additionally control for literacy. Overall, it seems that once migrants have the permission to work, they face no further disadvantages if they do not possess an EU passport.

6.3. *Heterogeneity across Individual Characteristics*

The extent to which observed wage gaps are associated with differences in literacy may depend on individual characteristics. We explore this possibility in Table 5 by restricting the estimation sample to full-time employees, men, women, and West German residents, respectively (columns 4–15). The raw wage gap is smallest for women, and the largest gap after including education and literacy persists in the male sample—although it is not statistically significantly different from zero.

Finally, in order to account for the idea that language learning is easier at young ages and is facilitated by time spent in the school system of the destination country (see, e.g., Chiswick and Miller, 2008; Stevens, 2004; Wiley *et al.*, 2004), we also construct samples which compare natives to migrants who were older than 11 or younger than 12 at the time of migration (columns 16–21). Here too, the usual pattern of results is observed, but—as expected—the raw wage gap is smaller for those with a lower age at migration.

Overall, these specifications underscore the robustness of our results. After controlling for literacy, none of the additional estimates suggest a statistically significant relation between the migrant indicator and wages.

7. Conclusion

This paper uses newly available information on literacy from the German "leo.—Level-One Studie" (LEO) and investigates whether the employment and wage gap between natives and migrants is related to the potentially lower language proficiency of migrants. The LEO dataset includes results of practical reading and writing tests, which minimize the measurement error usually introduced by self-reported items of language proficiency in other surveys. Another advantage of the data is that the literacy tests are conducted in the same way for migrants and natives, and therefore LEO supplies a measure of language skills that is readily comparable between the two groups. This enables us to directly and reliably investigate the relationship between literacy differences and their impact on migrant–native employment and the wage gap.

Interaction terms between the migrant and literacy variables show that the relationship between literacy and employment/wages is the same for migrants and natives, which suggests that the test scores are not differentially affected by other unobserved productivity relevant skills. We find that a one standard deviation increase in the literacy score is associated with a 3.9 percentage points higher probability of being employed, and with 7.2 percent higher wages. We control for education and linguistic distance in order to reduce bias due to confoundedness with ability and cultural differences, but do not claim to cleanly identify the causal effect of literacy. Identification of this effect is not strictly necessary for our central result: the migrant–native employment and wage gaps disappear completely and become insignificant when literacy levels are taken into account—that is, the differences in the labor market outcomes are fully explained by the literacy gap. Sensitivity tests using different samples quantitatively and qualitatively back our results.

One important implication of our paper is that observed raw wage differentials are not necessarily related to discrimination against migrants on the German labor market, because literacy is relevant to productivity in and of itself and can also be complementary to other forms of human capital. An implication in terms of public policy could be to specifically aim at increasing the reading and writing abilities of migrants in order to improve their economic position. This could also prove crucial when it comes to labor market integration of the large numbers of refugees who are currently arriving in Europe and do not speak the language of their host country—as Chiswick (2016, p. 31) puts it: "The economic success of migrants depends heavily on how well and quickly they learn the language of their new country."

References

Adams, R. and M. Wu, *PISA 2000 Technical Report*, Organization for Economic Co-operation and Development (OECD), Paris, 2002.

Aldashev, A., J. Gernandt and S. L. Thomsen, "The Immigrant–Native Wage Gap in Germany," *Journal of Economics and Statistics (Jahrbuecher fuer Nationaloekonomie und Statistik)*, 232, 490–517, 2012.

Algan, Y., C. Dustmann, A. Glitz, and A. Manning, "The Economic Situation of First and Second-Generation Immigrants in France, Germany and the United Kingdom," *Economic Journal*, 120, F4–F30, 2010.

Azariadis, C. and A. Drazen, "Threshold Externalities in Economic Development," *Quarterly Journal of Economics*, 105, 501–26, 1990.

Bakker, D., A. Müller, V. Velupillai, S. Wichmann, C. H. Brown, P. Brown, D. Egorov, R. Mailhammer, A. Grant, and E. W. Holman, "Adding Typology to Lexicostatistics: A Combined Approach to Language Classification, *Linguistic Typology*, 13, 169–81, 2009.

Bartolucci, C., "Understanding the Native–Immigrant Wage-Gap Using Matched Employed Employee: Evidence from Germany," *Industrial and Labor Relations Review*, 64, 1166–202, 2014.

Bell, B. D., "The Performance of Immigrants in the United Kingdom: Evidence from the GHS," *Economic Journal*, 107, 333–44, 1997.

Bilger, F., R. Jäckle, B. von Rosenbladt, and A. Strauss, "Studiendesign, Durchführung und Methodik der leo.—Level-One Studie," in A. Grotlüschen and W. Riekmann (eds), *Funktionaler Analphabetismus in Deutschland, Ergebnisse der ersten leo.—Level-One Studie*, Waxmann Verlag, Münster, 77–105, 2012.
Bleakley, H. and A. Chin, "Language Skills and Earnings: Evidence from Childhood Immigrants," *Review of Economics and Statistics*, 86, 481–96, 2004.
Borjas, G. J., "Assimilation, Changes in Cohort Quality, and the Earnings of Immigrants," *Journal of Labor Economics*, 3, 463–89, 1985.
Carliner, G., "Wage Differences by Language Group and the Market for Language Skills in Canada," *Journal of Human Resources*, 16, 384–99, 1981.
Chiswick, B. R., "The Effect of Americanization on the Earnings of Foreign-Born Men, *Journal of Political Economy*, 96, 897–922, 1978.
———, "The Economic Progress of Immigrants: Some Apparently Universal Patterns," in W. Fellner (ed.), *Contemporary Economic Problems*, American Enterprise Institute, Washington, D.C., 357–99, 1979.
———, "The Earnings of White and Coloured Male Immigrants in Britain," *Economica*, 47, 81–7, 1980.
———, "Speaking, Reading, and Earnings among Low-Skilled Immigrants," *Journal of Labor Economics*, 9, 149–70, 1991.
———, "Language in the Immigrant Labor Market: The Immigrant Experience in Canada and the United States," in B. R. Chiswick (ed.), *Immigration, Language and Ethnicity: Canada and the United States*, American Enterprise Institute, Washington, D.C., 1992.
———, "Tongue Tide: The Economics of Language Offers Important Lessons for How Europe Can Best Integrate Migrants," *IMF Finance & Development*, 53, 31–3, 2016.
Chiswick, B. R. and P. W. Miller, "The Endogeneity between Language and Earnings: International Analyses," *Journal of Labor Economics*, 13, 246–88, 1995.
———, "Linguistic Distance: A Quantitative Measure of the Distance between English and Other Languages," *Journal of Multilingual and Multicultural Development*, 26, 1–11, 2005.
———, "A Test of the Critical Period Hypothesis for Language Learning," *Journal of Multilingual and Multicultural Development*, 29, 16–29, 2008.
———, "Negative and Positive Assimilation, Skill Transferability, and Linguistic Distance, *Journal of Human Capital*, 6, 35–55, 2012.
———, "International Migration and the Economics of Language," in B. R. Chiswick and P. W. Miller (eds), *Handbook of the Economics of International Immigration*, vol. 1A, Elsevier, Amsterdam, 211–69, 2015.
Coulombe, S. and J.-F. Tremblay, "Literacy and Growth," *B.E. Journal of Macroeconomics*, 6, 1–34, 2006.
Coulombe, S., J.-F. Tremblay, and S. Marchand, "Literacy Scores, Human Capital and Growth across Fourteen OECD Countries," Statistics Canada, Ottawa, 2004.
Denny, K., C. Harmon, and M. Roche, "The Distribution of Discrimination in Immigrant Earnings-Evidence from Britain 1974–1993," IFS Working Papers W97/19, 1997.
Dustmann, C., "Earnings Adjustment of Temporary Migrants," *Journal of Population Economics*, 6, 153–68, 1993.
———, "Speaking Fluency, Writing Fluency and Earnings of Migrants," *Journal of Population Economics*, 7, 133–56, 1994.
Dustmann, C. and F. Fabbri, "Language Proficiency and Labour Market Performance of Immigrants in the UK," *Economic Journal*, 113, 695–717, 2003.
Dustmann, C. and A. Glitz, "Migration and Education," in E. A. Hanushek, S. Machin, and L. Wössmann (eds), *Handbook of the Economics of Education*, vol. 4, Elsevier, Amsterdam, 327–439, 2011.
Dustmann, C. and A. van Soest, "Language Fluency and Earnings: Estimation with Misclassified Language Indicators, *Review of Economics and Statistics*, 83, 663–74, 2001.
———, "Language and the Earnings of Immigrants," *Industrial and Labor Relations Review*, 55, 473–92, 2002.
Dustmann, C., T. Frattini, and I. P. Preston, "The Effect of Immigration along the Distribution of Wages," *Review of Economic Studies*, 80, 145–73, 2013.
Ferrer, A., D. A. Green, and W. C. Riddell, "The Effect of Literacy on Immigrant Earnings," *Journal of Human Resources*, 41, 380–410, 2006.
Finnie, R. and R. Meng, "Literacy and Labour Market Outcomes: Self-Assessment versus Test Score Measures," *Applied Economics*, 37, 1935–51, 2005.
Friedberg, R. M., "The Impact of Mass Migration on the Israeli Labor Market," *Quarterly Journal of Economics*, 116, 1373–408, 2001.

Green, D. A. and C. Riddell, "Literacy and Earnings: An Investigation of the Interaction of Cognitive and Unobserved Skills in Earnings Generation," *Labour Economics*, 10, 165–84, 2003.
Grotlüschen, A. and W. Riekmann, "leo.—Level One Study, Literacy of Adults at the Lower Rungs of the Ladder," Press brochure, University of Hamburg, http://blogs.epb.uni-hamburg.de/leo, 2011.
Hanushek, E. A. and L. Woessmann, "The Role of Cognitive Skills in Economic Development," *Journal of Economic Literature*, 46, 607–8, 2008.
Hartig, J. and W. Riekmann, "Bestimmung der Level-Grenzen in der leo.—Level-One Studie," in A. Grotlüschen and W. Riekmann (eds), *Funktionaler Analphabetismus in Deutschland, Ergebnisse der ersten leo.—Level-One Studie*, Waxmann Verlag, Münster, 106–21, 2012.
Isphording, I. E., "Disadvantages of Linguistic Origin—Evidence from Immigrant Literacy Scores," *Economics Letters*, 123, 236–9, 2014.
Isphording, I. E. and S. Otten, "Linguistic Barriers in the Destination Language Acquisition of Immigrants," *Journal of Economic Behavior and Organization*, 105, 30–50, 2014.
Jäckle, R. and O. Himmler, "Health and Wages: Panel Data Estimates Considering Selection and Endogenity," *Journal of Human Resources*, 45, 364–406, 2010.
Junker, B., L. S. Schofield, and L. J. Taylor, "The Use of Cognitive Ability Measures as Explanatory Variables in Regression Analysis," *IZA Journal of Labor Economics*, 1:4, 2012.
Kaas, L. and C. Manger, "Ethnic Discrimination in Germany's Labour Market: A Field Experiment," *German Economic Review*, 13, 1–20, 2012.
McIntosh, S. and A. Vignoles, "Measuring and Assessing the Impact of Basic Skills on Labour Market Outcomes," *Oxford Economic Papers*, 53, 453–81, 2001.
Mincer, J., "Investment in Human Capital and Personal Income Distribution," *Journal of Political Economy*, 66, 281–302, 1958.
———, *Schooling, Experience, and Earnings*, 3rd edn, National Bureau of Economic Research, New York, 1974.
Miranda, A. and Y. Zhu, "English Deficiency and the Native–Immigrant Wage Gap," *Economics Letters*, 118, 38–41, 2013.
Ozden, C., I. C. Neagu, and A. Mattoo, *Brain Waste? Educated Immigrants in the U.S. Labor Market*, World Bank, Washington, D.C., 2005.
Pischke, J.-S., "Assimilation and the Earnings of Guestworkers in Germany," ZEW Discussion Papers, 92-17, 1992.
Rubin, D. B., *Multiple Imputations for Non-response in Surveys*, Wiley, New York, 1987.
Sanders, C. and R. Lessem, *The Native–Immigrant Wage Gap in the United States*, 2013 Meeting Papers, Society for Economic Dynamics, 2013.
Sprietsma, M., "Discrimination in Grading: Experimental Evidence from Primary School Teachers," *Empirical Economics*, 45, 523–38, 2013.
Statistisches Bundesamt, *Bevölkerung und Erwerbstätigkeit—Bevölkerung mit Migrationshintergrund—Ergebnisse des Mikrozensus 2009*, Fachserie 1, Reihe 2.2, Statistisches Bundesamt, Wiesbaden, 2011.
Stevens, G., "Using Census Data to Test the Critical-Period Hypothesis for Second-Language Acquisition," *Psychological Science*, 15, 215–16, 2004.
Vignoles, A., A. De Coulon, and O. Marcenaro-Gutierrez, "The Value of Basic Skills in the British Labour Market, *Oxford Economic Papers*, 63, 27–48, 2010.
Wichmann, S., E. W. Holman, and C. H. Brown (eds), The ASJP Database (version 17), 2016.
Wiley, E., E. Bialystok, and K. Hakuta, "New Approaches to Using Census Data to Test the Critical-Period Hypothesis for Second-Language Acquisition," *Psychological Science*, 16, 341–3, 2004.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Table A.1: Description of Variables
Table A.2: Summary Statistics (Weighted)
Table A.3: Migrants, Different Definitions (I), Labor Force and Employees (in Parentheses)
Table A.4: Migrants, Different Definitions (II), Labor Force and Employees (in Parentheses)

Table A.5: Wage equation, Consistency Checks. Dependent variable: Log Wages. Language Based Migrant Definition
Figure A.1: Simulation of Wage Gap at Different Returns to Literacy
Figure A.2: Distribution of Literacy, Labor Force (Aged 18–64)
Figure A.3: Distribution of Literacy Scores, Labor Forche (Aged 18–64)
Figure A.4: LEO-Test Items, α-levels 1 and 2
Figure A.5: LEO-Test Items, α-level 3 and 4