MPRA

Munich Personal RePEc Archive

Literacy and the Migrant-Native Wage Gap

Oliver Himmler and Robert Jaeckle

Max Planck Institute Bonn, TH Nuernberg

September 2014

Literacy and the Migrant–Native Wage Gap[*]

Oliver Himmler[‡]

Max Planck Institute, Bonn

Robert Jäckle

TH Nürnberg Georg Simon Ohm

this version: September 2014

Abstract

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 study), 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.

Keywords: Literacy, Migration, Employment, Earnings, Wage Gap, Discrimination

JEL Classification: J24, J31, J61

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 (Chiswick 1978, Borjas 1985, Sanders and Lessem 2013), Canada (Ferrer, Green, and Riddell 2006), the UK (Chiswick 1980, Denny, Harmon, and Roche 1997, Bell 1997, Miranda and Zhu 2013), and Germany (Pischke 1992, Dustmann 1993, Aldashev, Gernandt, and Thomsen 2012, Bartolucci 2009), among others. Non-trivial differences often remain even after factoring out determinants of wages such as educational attainment.[1]

It is important to investigate whether such observed inequalities can be attributed to migrants actually being discriminated against by employers or whether they simply lack competencies or qualifications 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 one resides in. 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 e.g. on how to keep up 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 doesn't 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, e.g. if they cannot find the right job because they are unable to read job offers. While this complementarity theoretically also matters for natives, it is far less likely to

---

[1]Algan, Dustmann, Glitz, and Manning (2010) show that, unconditionally, in the UK migrants earn higher wages on average than natives. At the same time migrants have higher schooling on average, and conditional on education their earnings fall behind those of natives.

observe a native with a high stock of human capital who doesn't speak the native language than it is to observe a migrant for whom this is true. In fact, in the data that we use, only 4% of natives with a university degree are functionally illiterate in the German language, whereas this applies to 24% 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, Neagu, and Mattoo 2005, Dustmann, Frattini, and Preston 2013), and therefore substantially contribute to wage differentials with equally educated natives.

Assessing 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, Chiswick 1992, Chiswick and Miller 1995, Chiswick and Miller 2014 provide a concise survey). Prominent examples include the US census or the German Socioeconomic Panel (GSOEP). These assessments are of course subject to measurement error, potentially also systematic bias because self assessments depend on the individual reference group, i.e. 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 contains 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 circumvents 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 even larger than any upward bias caused by literacy being confounded with innate ability or motivation (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 don't have a good command of German, and therefore the wage gap will to a large extent be identified off variation in this literacy region.

The data that we use is especially well suited for this purpose and to our knowledge it is novel to the economics literature. It stems from the Level One Study (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, the Level One. Upon its release, LEO gained quite some media attention, mainly because it uncovered that the prevalence of illiteracy is roughly twice as high as previously thought. Some 8400 individuals were interviewed, representative of the German population, and 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% of natives and 76% 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 the literature on the general effects of literacy on wages needs top cope with. 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 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 is 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 it is captured 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 are therefore confident that the literacy variable does not measure different things for the two groups of individuals.

This is the backdrop against which we will attempt to 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 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 is due to differences in literacy. The initial differential in employment is roughly 6% to the disadvantage of migrants, and considering differences in education cuts this gap in half. Migrants who have spent more time in Germany are 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 proxy for cultural distance. 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: adding education cuts the initial 14% 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% lower earnings than those who reach at least LEO level four. Accordingly, when correcting for the fact that migrants have lower proficiency in German, no wage gap remains. These results indicate 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. This result is robust across a number of subsamples where the results also show the same pattern, except for the subgroup of migrants who arrived in Germany before the age of 12. Among these individuals there is no raw wage differential to begin with.

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 (2014) 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 GSOEP which provides self assessments of migrant language skills, and shows that higher levels of literacy go with higher earnings of migrants. Dustmann and van Soest (2001) revisit the topic, and show that the self assessed language skills from the GSOEP 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, Gernandt, and Thomsen (2012) also rely on GSOEP data, and show that increased language skills go 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 higher 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, Tremblay, and Marchand (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 (Vignoles, De Coulon, and Marcenaro-Gutierrez 2010, McIntosh and Vignoles 2001, Green and Riddell 2003. 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 recent 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, Sprietsma (2013) in another field

experiment finds evidence that 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. Section 6 concludes.

2 Data and Descriptives

The data used in our analysis is taken from the LEO–LEVEL-ONE STUDY provided by the University of Hamburg (see Grotlüschen and Riekmann 2011). LEO is representative for 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.[2] LEO conducts practical reading and writing tests as part of the face-to-face interview which enable us to cope with the substantial measurement error introduced by the self-reported language proficiency scales that many studies use (Carliner 1981; Chiswick 1991; see Dustmann and Glitz 2011 for a comprehensive survey). These competence tests allow for a categorization of the sample population into five groups: 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 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*, i. e. they are able to read and write single sentences but fail even with short texts (see figures 5 and 6 for sample test items). Respondents *below grade school level* (= α-level 4) cannot read or write texts at a level that is expected at the end of 4th grade – these people typically avoid reading and writing, even with texts that include commonly known words only. 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 (= the Level One),

---

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

the LEO-study provides a novel dataset and fills a gap in the existing literature.[3] LEO includes 8,436 observations from all German states (Bundesländer) and consists of two sub–samples. 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 sufficiently read and write.[4] Combining the two sub-samples 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 sub-samples and extract data on the variables described in tables 5 and 6 in the appendix. As recommended in Bilger et. al. (2012), we drop 20 observations with obviously invalid information on the literacy variables. Also excluded are 44 units who can not be uniquely assigned to one educational level. We further constrain the sample to individuals who belong to the labor force, i. e. those who are engaged in full-time (3,107 obs.) or part-time work (1,418 obs.) and individuals who report to be currently unemployed (1,126 obs.). After these steps we obtain a final unweighted estimation sample of 5,651 observations including 568 migrants.

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

2.1 Migration, Wage, and Literacy variables

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. German statistics typically use the category of "individuals with migration background": it includes everyone who was born to at least one parent that is non-German or that migrated to Germany. Since we are interested in the relation between language proficiency and labor market

---

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

[4]For further details see Bilger et. al. (2012).
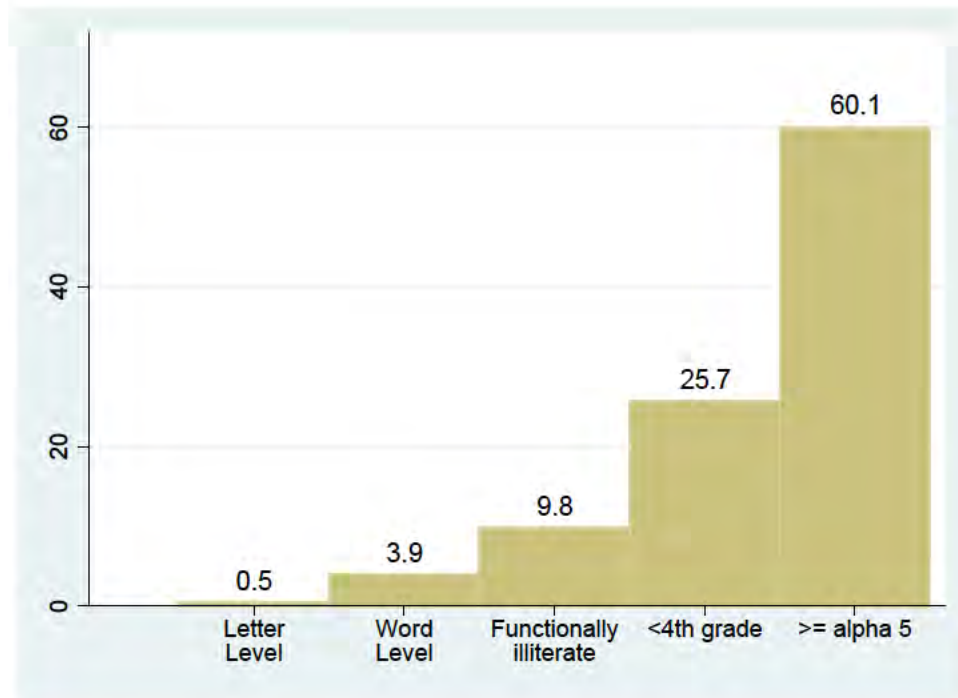
Source: LEO–LEVEL-ONE STUDY, 2010, own calculations. Note: Averaged shares (probability weighted) based on five plausible values. 5,651 observations including 568 migrants.

Figure 1: **Distribution of Literacy, Labor Force (aged 18–64)**

outcomes – not citizenship effects or born abroad effects per se – we use another definition: we label as migrants all individuals who do not report their native tongue to be German.[5] In comparison to "migration background" this excludes those with a migration background who name German as their first language – they are assigned to the native group. Even though the definitions produce largely overlapping groups, ours tends to create larger raw wage gaps. The reason is that among individuals with a "migration background" those who name German as their mother tongue fare better on the labor market than those who have a foreign mother tongue. Given our definition, 13.6% of the labor force and 12.8% of the employed qualify as migrants (see table 6 in the appendix).

---

[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".

Literacy variables and plausible values. Instead of a single cognitive literacy score, the LEO-data-set 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.[6] When using plausible values, measurement error in the literacy scores is negligible (see Junker, Schofield, and Taylor 2012) and the efficiency of population estimates improves. However, as each draw of a plausible value includes a random error component, these values can not 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 1). 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%.[7]

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 wages is an intrusive question to many respondents. In the LEO-survey 37% of the income data is missing. However, 87% of those refusing to quote their salary were willing to classify it within certain ranges ($\leq$ € 400, 401–1000 € and $>$ € 1000). For those who only provided a class of income, we impute predicted gross wages based on linear wage regressions using respondents in the respective class who provided an exact income. Predictions are based on age, sex, education, occupation, working hours and region.

2.2 Descriptive comparison of the migrant and native labor force

Based on our definition of migrant status, in 2010 about 16% of the migrant and around 9% of the native labor force were unemployed (see figure 2 and table 6). On average, monthly gross wages of natives were about € 366 higher than those of migrants, and we find virtually no difference in hours worked per week.

---

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

[7]Because we have five plausible values, a single individual can be allocated to different α-levels in different draws.
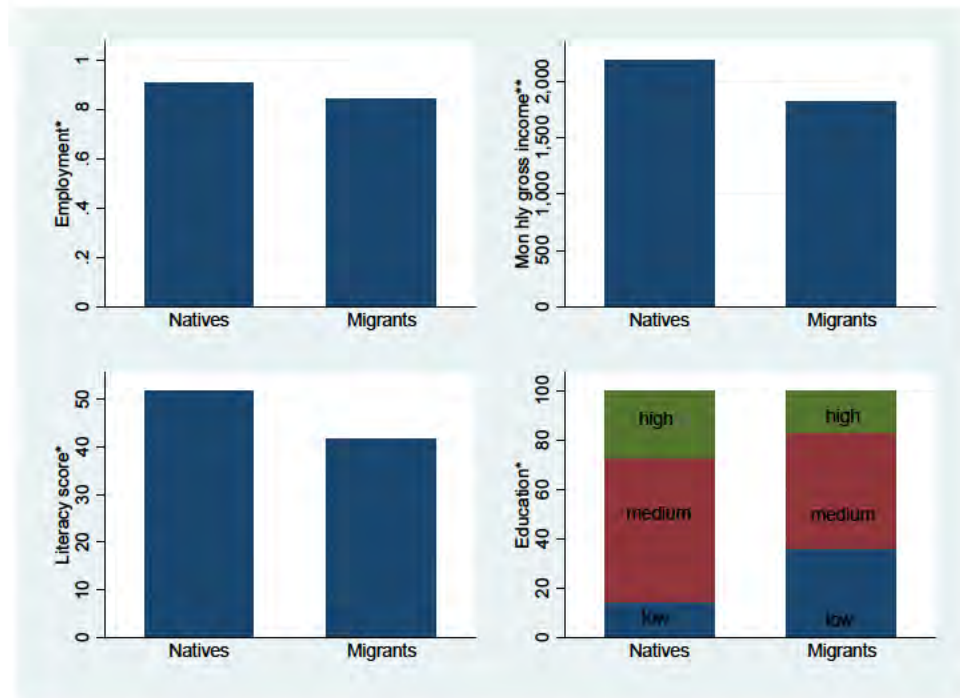
Source: LEO–LEVEL-ONE STUDY, 2010, own calculations. Note: Averaged shares (probability weighted) based on five plausible values. *5,651 observations including 568 migrants. **4,525 observations including 427 migrants.

Figure 2: **Migrant and Native Labor Force, important differences**

Figure 3 shows that the literacy skills of the migrant and native workforce 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 2 and table 6).[8]

Using discrete literacy levels, we find that strict illiteracy (functional illiteracy) affects 17.6% (25.9%) of the migrants, compared to a mere 2.4% (7.3%) among Germans (see figure 4 and table 6). Furthermore, 32.6% (24.6%) of the migrants (natives) cannot read or write texts at a level that is expected at the end of 4th

---

[8]The hypothesis of the equality of means between migrants and natives can be rejected on the basis of a standard t-tests (p-value = 0). Additionally, we conduct a Kolmogorov-Smirnov two-sample test, which clearly rejects the hypothesis of the equality of literacy distributions between the two groups (p-value = 0).

Figure 3: **Distribution of Literacy Scores, Labor Force**

grade (α-level 4). Overall, the LEO-survey reveals that the share of people who have adequate reading and writing skills among the migrant work force (23.8%) is less than half the share among natives (65.8%). Moreover, the proportion of migrants who attain a test score in the range of α-level 3 or below (= cumulative functional illiteracy, 43.5%) is more than four times as high as the respective share among natives (9.7%).

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 2 provides a glance at educational differences between migrants and natives.[9] The numbers suggest that, on average,

---

[9]The education variable combines highest school and occupational degrees and generally distinguishes three levels—low ($\leq$ lower secondary education, occupational training of no more than one year), medium (= upper secondary education or post-secondary non-tertiary education, three
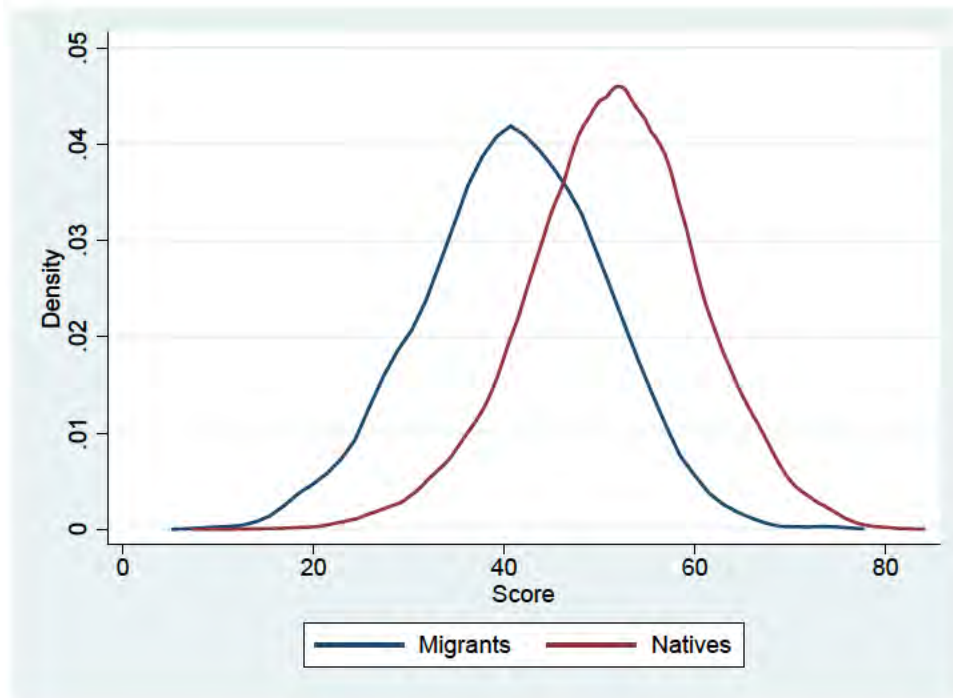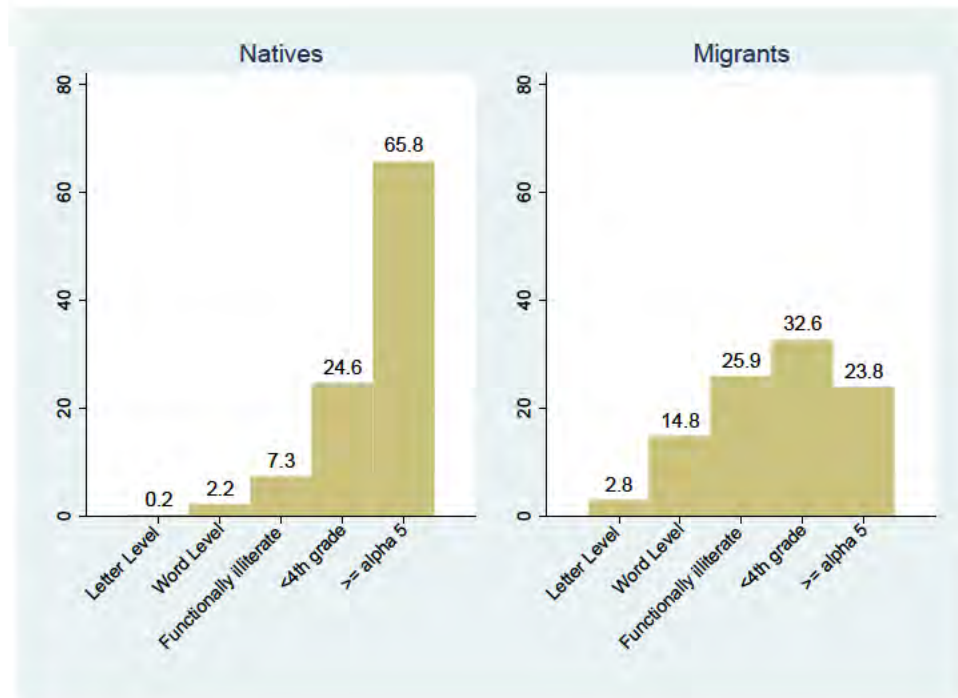
Source: LEO–LEVEL-ONE STUDY, 2010, own calculations. Note: Averaged shares (probability weighted) based on five plausible values. 5,651 observations including 568 migrants.

Figure 4: α-Levels, Labor Force

migrants indeed possess lower educational degrees than natives: 36% (14.3%) of the migrant (native) population has at most a lower educational or occupational degree, 47.2% (58.2%) has a medium educational or occupational qualification, and 16.8% (27.5%) are highly qualified. Almost half of the immigrants (44.5%) 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 grow-

---

year occupational degree) and high education (= bachelor's degree or higher, master craftsman degree).

ing up. Migrants, on the other hand, acquire proficiency in the host country's language at high costs.[10] These costs differ according to e. g. their educational background or the distance of the mother tongue to the destination language. While it seems obvious that migrants on average have lower literacy skills in the destination country language, it is still informative to see to what extent the groups differ in literacy, and what the determinants of literacy are.

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.). We estimate each specification five times—once with each relevant plausible value variable—, 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, a gender dummy, a dichotomous variable for having a partner, 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) shows that the average migrant's language command 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%. 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 down the functional illiteracy gap to 27 percentage points (for newly arrived migrants these gaps are of course much larger, because time since migration is centered at its mean of roughly 22 years).

The degree of difficulty in acquiring the host country's language varies depending on the migrant's first language.[11] 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 using 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

---

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

[11]See e.g. Chiswick and Miller (2005), (2012), (2014), and Isphording and Otten (2014).

Table 1: Literacy Gap, Labor Force. Dependent variable: Literacy Score (L. S.) / Functional Illiteracy (≤ Funct.)

VARIABLES	(1) L. S.	(2) ≤ Funct.	(3) L. S.	(4) ≤ Funct.	(5) L. S.	(6) ≤ Funct.	(7) L. S.	(8) ≤ Funct.
Migrant	-9.558*** (0.626)	0.305*** (0.036)	-8.495*** (0.625)	0.273*** (0.036)			-8.526*** (0.618)	0.274*** (0.036)
Afro-asiatic language					-9.425*** (2.214)	0.378*** (0.118)		
Altaic language					-8.942*** (1.133)	0.312*** (0.078)		
Germanic language					-5.550** (2.391)	0.127 (0.110)		
Iranian language					-12.785*** (1.889)	0.466*** (0.098)		
Romanic languages					-8.213*** (1.712)	0.297*** (0.107)		
Slavic language					-7.703*** (0.774)	0.214*** (0.045)		
Indo-european language					-9.955*** (2.278)	0.339*** (0.123)		
Other language group					-10.246*** (2.165)	0.305** (0.139)		
Centered (Cntrd.) ling. distance × migrant							-0.089 (0.073)	0.005* (0.003)
Education medium			4.284*** (0.404)	-0.156*** (0.022)	4.183*** (0.401)	-0.149*** (0.021)	4.278*** (0.403)	-0.155*** (0.022)
Education high			7.891*** (0.493)	-0.218*** (0.024)	7.724*** (0.493)	-0.209*** (0.023)	7.844*** (0.497)	-0.215*** (0.024)
Cntrd. years since migration × migrant	0.118*** (0.041)	-0.004** (0.002)	0.122*** (0.040)	-0.004** (0.002)	0.117*** (0.042)	-0.005** (0.002)	0.122*** (0.040)	-0.004** (0.002)
Male	-2.879*** (0.368)	0.065*** (0.012)	-3.186*** (0.371)	0.073*** (0.012)	-3.151*** (0.375)	0.069*** (0.012)	-3.179*** (0.371)	0.072*** (0.012)
Has partner	1.121** (0.444)	-0.034** (0.017)	0.844** (0.409)	-0.027* (0.016)	0.806* (0.412)	-0.026* (0.016)	0.834** (0.411)	-0.027* (0.016)
No. children ≤ 6 years	-0.051 (0.429)	0.008 (0.013)	0.050 (0.419)	0.004 (0.013)	0.093 (0.414)	0.001 (0.013)	0.061 (0.419)	0.003 (0.013)
No. children. 7-13 years	0.104 (0.276)	0.010 (0.014)	0.256 (0.267)	0.006 (0.014)	0.283 (0.264)	0.004 (0.013)	0.266 (0.266)	0.005 (0.014)
No. children 14-17 years	-0.473 (0.475)	0.013 (0.015)	-0.344 (0.451)	0.010 (0.014)	-0.323 (0.445)	0.008 (0.015)	-0.323 (0.446)	0.008 (0.014)
Constant	55.248*** (4.064)	-0.061 (0.125)	49.398*** (3.960)	0.126 (0.126)	49.687*** (4.082)	0.119 (0.127)	49.551*** (3.941)	0.118 (0.125)
Cohort f. e. χ^2_{46}	59.16*	80.17***	91.25***	92.18***	89.81***	93.14***	91.52***	93.16***
Population size f. e. χ^2_6	1.022	0.343	1.276	0.425	1.489	0.409	1.367	0.476
County f. e. χ^2_{83}	2,031***	1,170***	1,916***	1,199***	1,845***	1,045***	1,902***	1,129***
Interviewer f. e. χ^2_{281}	1.647×10⁶***	6.147×10⁶***	1.075×10⁶***	4.613×10⁶***	3.644×10⁶***	5.217×10⁶***	1.676×10⁶***	3.490×10⁶***

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 compare pairs of languages (see Bakker et. al. 2009).[12] Ceteris paribus, the literacy gap is smallest for migrants with Germanic (0.56 std. dev. and 13 percentage points), Slavic (0.78 std. dev. and 21 percentage points), and Romanic language background (0.83 std. dev. and 30 percentage points). The largest difference is found with respect to the Iranian language tree (1.3 std. dev. 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 0.5 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) between 1.23% and 2.21% (1.31% and 1.47%). Being male is associated with poorer reading and writing abilities, 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 e.g. 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. 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, also for

---

[12]For more details see Isphording and Otten (2014). They were among the first who used the newly available distance measure for Germany and studied its effect on (self-assessed) language fluency of migrants using the German Social Economic Panel.

low skilled jobs it is important to have a sufficient command of the host country's language, e.g. to follow work instructions or to comply with the health and safety regulations of the employer.

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 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 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. 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) to (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%). Correcting for the fact that migrants and natives on average 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) to (6) virtually eliminates the migrant-native employment gap—i.e. 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.[13]

In columns (3) and (4) we restrict the relationship between literacy and employment to be the same for migrants and natives. An increase of the literacy score by one standard deviation increases the probability of being employed by 3.9 percentage points (column (3)). 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

---

[13]When interpreting the results, 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 to some extent explain the results. This is true for both migrants and natives and so should not matter much for our assessment of the employment gap.

Table 2: EMPLOYMENT GAP, ALL RESPONDENTS. 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)
<4th grade level literacy				-0.037***		-0.036***
				(0.012)		(0.013)
Std. Lit. Score × migrant					0.002	
					(0.017)	
Centered Alpha1/2 × migrant						0.043
						(0.072)
Centered Functional × migrant						0.046
						(0.053)
Centered <4th 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 ling. dist. × 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)
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)
No. 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)
No. 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)
No. 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 f. e. χ^2_{46}	108.5***	130.6***	94.48***	94.23***	93.70***	94.17***
Population size f. e. χ^2_6	18.38***	18.90***	19.09***	18.48	19.09***	18.36***
County f. e. χ^2_{81}	2.1×10^7***	3.03×10^7 ***	26,13***	22,94***	23,71***	20,46***
Interviewer f. e.χ^2_{286}	842,5***	3.12×10^6***	1.59×10^8***	2.03×10^8***	1.24×10^8***	5.51×10^7***

Source: LEO–LEVEL-ONE STUDY , 2010, own calculations. *Note:* 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 control for interview duration and for cohort, population size, county and interviewer fixed effects. Standard errors in parentheses clustered by counties and adjusted for variation between the five estimates, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

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 measure the same thing for both migrants and natives, and is not differentially confounded with motivation or ability.

We further control for the linguistic distance between home and host country in specifications (3) to (6) and find that higher distance significantly reduces the probability to work. Since we are already holding literacy constant, the distance coefficient cannot be driven by different language skills. Rather, it seems plausible that linguistic distance captures differences in culture, because countries with similar languages may also be more similar culturally.

The coefficients of the number of 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 five percentage points higher probability of being employed. Having a partner increases the probability to work 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 to work, whereas older children (aged between 7 and 17) do not seem to make a difference. The education parameters are significantly positive, and as expected the magnitude increases in education level. Overall, our results for the control variables support the results of other studies, for example, in Britain (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, informational deficiencies with respect to the host country's labor market, but of course also by poor command of the host country's 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 log of working hours per week, and two dummy variables indicating whether the individual is part-time or 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. Following Bleakley and Chin (2004) we add linguistic distance, which captures cultural differences and via this channel may also control for factors such as missing networks, and informational deficiencies with respect to the host country's labor market. Both variables are centered at the mean and set to zero for natives. Making use of the plausible values in the LEO data set we estimate each specification five times to reduce measurement error in the language command of migrants and natives to a minimum.

All standard controls have the expected signs and are estimated to be statistically significant at the 1% 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 linguistic distance are positively correlated with wages but are not statistically significant. As linguistic distance already captures differences in cultural dimensions as well as network effects, and informational deficiencies related to the host country's labor market 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% (column 3). The results in column (4) show that wages of literates differ from individuals on α-level 1 or 2, α-level 3, and α-level 4 by 27%, 19.2%, and 7.4%. Allowing the literacy coefficients to vary between migrants and natives in columns (5) and (6) does not change any of the restricted 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 find that migrants earn on average 14.6% less then 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 literary variables in columns (3) to (6), however, the wage gap vanishes completely and is no longer statisti-

Table 3: Wage Equation, All respondents. 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)
<4th grade level literacy				-0.074***		-0.078***
				(0.024)		(0.027)
Std. Lit. Score × migrant					0.025	
					(0.032)	
Centered Alpha1/2 × migrant						0.073
						(0.144)
Centered Functional × migrant						-0.017
						(0.103)
Centered <4th 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 ling. dist. × 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)
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 f. e. χ^2_{46}	1,065***	729.1***	724.0***	680.2***	695.6***	629.9***
Population size f. e. χ^2_6	6.664	5.677	5.567	5.594	5.624	5.560
County f. e. χ^2_{80}	1.29×10^6***	1.79×10^6***	11,767***	9,478***	10,257***	10,497***
Interviewer f. e. 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. *Note:* Averaged parameters from five weighted OLS estimates based on the relevant plausible value. 4,525 observations including 427 migrants. All regressions additionally control for interview duration and for cohort, population size, county and interviewer fixed effects. Standard errors in parentheses clustered by counties and adjusted for variation between the five estimates, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 4: WAGE EQUATION, DIFFERENT SAMPLES. DEPENDENT VARIABLE: LOG WAGES.

VARIABLES	(1) All	(2) All	(3) All	(4) FT	(5) FT	(6) FT	(7) Male	(8) Male	(9) Male	(10) Fem	(11) Fem	(12) Fem
Migrant	-0.146***	-0.076***	-0.014	-0.184***	-0.100**	-0.049	-0.228***	-0.129***	-0.068	-0.075**	-0.038	0.023
	(0.029)	(0.028)	(0.030)	(0.048)	(0.047)	(0.044)	(0.041)	(0.039)	(0.042)	(0.032)	(0.040)	(0.047)
Std. Lit. Score			0.072***			0.063***			0.074***			0.066**
			(0.016)			(0.017)			(0.024)			(0.028)
Educ. 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)		(0.030)	(0.028)
Educ. 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)		(0.032)	(0.037)
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
Obs. migrants	427	427	427	291	291	291	237	237	237	190	190	190

VARIABLES	(13) West	(14) West	(15) West	(16) AgeMg>11	(17) AgeMg>11	(18) AgeMg>11	(19) AgeMg≤11	(20) AgeMg≤11	(21) AgeMg≤11
Migrant	-0.184***	-0.095***	-0.032	-0.184***	-0.075	-0.005	-0.008	0.046	0.081
	(0.020)	(0.034)	(0.025)	(0.049)	(0.057)	(0.054)	(0.049)	(0.067)	(0.074)
Std. Lit. Score			0.070***			0.068***			0.068***
			(0.017)			(0.016)			(0.015)
Educ. medium		0.244***	0.211***		0.229***	0.196***		0.216***	0.183***
		(0.022)	(0.024)		(0.024)	(0.027)		(0.024)	(0.026)
Educ. high		0.584***	0.526***		0.577***	0.519***		0.564***	0.508***
		(0.036)	(0.038)		(0.029)	(0.034)		(0.029)	(0.032)
Observations	3,488	3,488	3,488	4,393	4,393	4,393	4,230	4,230	4,230
Obs. migrants	395	395	395	295	295	295	132	132	132

Source: LEO−LEVEL-ONE STUDY, 2010, own calculations. *Note*: Averaged parameters from five weighted OLS estimates each one calculated using the relevant plausible value. All specifications include educations dummies, 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. Standard errors in parentheses clustered by counties and adjusted for variation between the five estimates, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
[a] Average number of observations—an individual can be allocated to different α-levels for different plausible values.

cally significant. This result is similar to what we find in Table 2 when analyzing employment probabilities. Thus, our results show that both the wage gap and employment differences between migrants and natives can be fully explained by the literacy gap.

Table 4 presents consistency checks of our results using alternative samples. We re-estimate columns (1)–(3) of table 3 and restrict the estimation sample to full time employees, men, women, and West German residents. Finally, in order to account for the idea that language learning outcomes worsen with age and increase with time spent in the destination country school system (see e.g. Chiswick and Miller 2008, Wiley, Bialystok, and Hakuta 2004, or Stevens 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 19–24). All specifications underscore the robustness of our results. None of the additional estimates suggest a significant relation between the migrant indicator and wages when we control for literacy. Though not significantly different from zero, the largest gap (6.8%) persists in the male sample. Interestingly, we find the sign of the wage gap to reverse in the female, and in the age at migration ≤ 11 sample suggesting that everything else equal migrants even earn more than natives in these populations.

6 Conclusions

This paper uses newly available information on literacy from the German LEO–LEVEL-ONE STUDY and investigates whether the employment and wage gap between natives and migrants is related to potentially lower language proficiency of migrants. The LEO data set 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.[14] This enables us to directly and reliably investigate the relationship between literacy differences and their impact on the migrant-native employment and 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 in-

---

[14]Usually surveys use filters to avoid asking natives about their linguistic abilities assuming that natives are fully capable of reading and writing their mother tongue.

crease in the literacy score is associated with a 3.9 percentage points higher probability of being employed, and with 7.2% 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. Identifying this effect is not strictly necessary for our central result: the migrant-native employment and wage gaps completely disappear and become insignificant when literacy levels are taken into account—i.e. the differences in 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 wage differentials measured in the data are not necessarily related to discrimination against migrants on the German labor market, because literacy is productivity relevant in and of itself and it also can be complementary to other forms of human capital. A policy implication could be to specifically aim at increasing the reading and writing abilities of migrants in order to improve their economic position. This recommendation is of course also true for natives with low linguistic abilities.

Appendix

Table 5: Description of Variables

Variable	Description
Migration	Variable indicating German (Migrant = 0) or foreign (Migrant = 1) mother tongue
Log wage	Log gross monthly wage
Literacy Score[a]	Continuous literacy score
Alpha 1 or 2[a]	Indicator for being able to read and write on the "letter" or "word level"
Functionally illiterate[a]	Indicator for being able to read and write single sentences but not short texts
<4th grade level literacy[a]	Indicator for being able to read and and write at 4th grade level
Education low	Lower secondary/second stage of basic education or below
Education medium	Upper secondary education or post-secondary non-tertiary education
Education high	First or second stage of tertiary education
Years since migration	Years since immigration to Germany
Ling. dist.	Levenshtein linguistic distance to German
Language groups	Language tree dummies (Altaic, Germanic, Romanic, Slavic, Indo-european, or other language)
Log. working hours	Log working hours per week
Male	Indicator for being male (male = 1) or female
Has partner	Individual has a partner / is married
No. of children	Number of children in three categories: 1) Up to 6 years old; 2) between 7 and 13 years old; 3) between 14 and 17 years old
Part time employed	Indicator for being part time (= 1) employed
Self employed	Indicator for being self (= 1) employed
Interview duration	Interview duration in minutes (including the competence tests)
Cohort f. e.	Indicators for birth cohorts
Population Size f. e.	Indicator for living in rural/urban region
County f. e.	Indicators for living in different German counties
Interviewer f. e.	Indicators for being interviewed by the same interviewer

[a]Five plausible values.

Table 6: Summary Statistics (weighted)

	Labor force			Employees		
	All (1)	Native (2)	Migrants (3)	All (4)	Native (5)	Migrants (6)
Migrant	0.136	0	1	0.128	0	1
	(0.343)	(0)	(0)	(0.334)	(0)	(0)
Employed	0.899	0.908	0.843	1	1	1
	(0.301)	(0.289)	(0.364)	(0)	(0)	(0)
Log. monthly wages	-	-	-	7.435	7.454	7.305
				(0.750)	(0.755)	(0.705)
Literacy Score[a]	49.96	51.36	41.02	50.57	51.86	41.72
	(9.856)	(9.140)	(9.560)	(9.618)	(8.932)	(9.488)
Alpha 1 or 2[a]	0.0446	0.0239	0.176	0.0361	0.0186	0.155
	(0.207)	(0.153)	(0.381)	(0.186)	(0.135)	(0.363)
Functionally illiterate[a]	0.0982	0.0728	0.259	0.0872	0.0637	0.248
	(0.298)	(0.260)	(0.439)	(0.282)	(0.244)	(0.432)
<4th grade level literacy	0.257	0.246	0.326	0.251	0.239	0.337
	(0.437)	(0.431)	(0.469)	(0.434)	(0.426)	(0.473)
Log. working hours	-	-	-	3.466	3.467	3.463
				(0.437)	(0.436)	(0.440)
Part time employed	-	-	-	0.293	0.290	0.312
				(0.455)	(0.454)	(0.464)
Self employed	-	-	-	0.135	0.135	0.135
				(0.342)	(0.342)	(0.342)
Age	42.67	43.07	40.10	42.96	43.34	40.31
	(10.98)	(10.99)	(10.57)	(10.77)	(10.75)	(10.51)
Education medium[b]	0.567	0.582	0.472	0.576	0.588	0.494
	(0.496)	(0.493)	(0.500)	(0.494)	(0.492)	(0.501)
Education high	0.260	0.275	0.168	0.277	0.292	0.172
	(0.439)	(0.446)	(0.375)	(0.448)	(0.455)	(0.378)
Years since migration	-	-	22.64	-	-	22.53
			(11.22)			(11.25)
Afro-asiatic language[c]	-	-	0.0547	-	-	0.0510
			(0.228)			(0.220)
Altaic language	-	-	0.240	-	-	0.243
			(0.428)			(0.429)
Iranian language	-	-	0.0467	-	-	0.0427
			(0.211)			(0.202)
Romanic languages	-	-	0.127	-	-	0.136
			(0.333)			(0.343)
Slavic language	-	-	0.355	-	-	0.341
			(0.479)			(0.474)
Indo-european language	-	-	0.0543	-	-	0.0536
			(0.227)		(0.226)	
Other language group	-	-	0.0627	-	-	0.0647
			(0.243)			(0.246)
Linguistic distance	-	-	93.54	-	-	93.12
			(7.03)			(7.6)
Male	0.544	0.533	0.610	0.539	0.529	0.608
	(0.498)	(0.499)	(0.488)	(0.499)	(0.499)	(0.489)
Has partner	0.748	0.746	0.761	0.776	0.775	0.782
	(0.434)	(0.435)	(0.427)	(0.417)	(0.418)	(0.413)
No. children ≤ 6 years	0.224	0.201	0.368	0.221	0.202	0.354
	(0.539)	(0.511)	(0.676)	(0.536)	(0.513)	(0.653)
No. children. 7-13 years	0.306	0.286	0.437	0.311	0.294	0.427
	(0.624)	(0.601)	(0.738)	(0.623)	(0.608)	(0.707)
No. children 14-17 years	0.181	0.176	0.216	0.186	0.183	0.205
	(0.451)	(0.443)	(0.498)	(0.454)	(0.450)	(0.480)
Interview duration[d]	1.179	1.173	1.238	1.242	1.242	1.242
	(18.41)	(19.33)	(5.568)	(20.43)	(21.3)8	(6.110)
Observations	5,651	5,083	568	4,525	4,098	427

Source: leo–Level-One Study, 2010, own calculations.
[a]Weighted means averaged over five plausible values drawn from the posterior distribution of the literacy score.
Standard deviations in parentheses. [b]Basis category: lower education.
[c]Basis category: Germanic language. [d]Interview duration is unweighted.

In Germany, the *leo.* items are the first items that record the scale of functional illiteracy which do not just fulfil the demands of the International Adult Literacy Survey (IALS), but surpass them in terms of theoretical basis.

The goals of *lea.* and *leo.* are also to differentiate at the lower end of the spectrum, roughly called Level One, between small increments and exactly described Alpha Levels.

27

Exercise type picture puzzles (Alpha Level 1 and 2)

Interviewer: Please look at the pictures and tick the right answer.

What has been damaged?	o pavement o sign o road

What is the ship called?	o Frau o Heidi o Hedi

Exercise type allocation (Alpha Level 2)

Interviewer: Now let's turn to the next page in the puzzle book. On the left you can see four titles and on the right four sentences about them. Please draw an arrow from the title to the sentence that goes with it. An arrow has already been drawn as an example.

Youth

A rich man and a poor man were standing about looking at one another. And the poor man says: If I wasn't poor, you wouldn't be rich.

Pain

How do you stay young? Live honourably, eat slowly and give the wrong age!

Death

Pain is only pleasant once it has abated.

Wealth
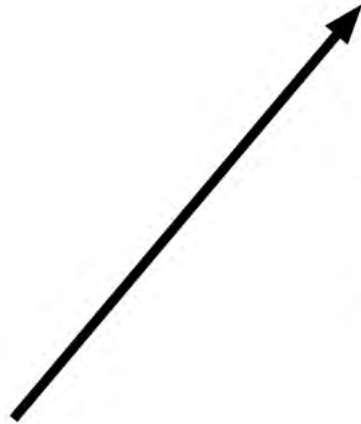
Dying is so hard and eternity is so long!

Source: LEO–LEVEL-ONE STUDY, 2010, Press Brochure, http://blogs.epb.uni-hamburg.de/leo/files/2011/12/leo-Press-brochure15-12-2011.pdf, 2011-12-15.

11

Figure 5: leo–Test Items, α-levels 1 and 2

Exercise type audio dictation (Alpha Level 4)

Interviewer: Some people say funny things at work. Please listen and note down the words the speaker states afterwards. (Note: the audio file is switched on if consent is given.)

One **baker** says to another: "Rumour has it that the beach is underneath the paving **stones** - have a look in the driveway".
"Doesn't matter" replies his apprentice "For my **holiday** I'm heading into the mountains anyway!"

Afterwards: dictation of the words baker, paving stones, driveway, holiday.

Figure 6: leo–Test Items, α-level 3 and 4

References

ADAMS, R., AND M. WU (EDITORS) (2002): "PISA 2000 Technical Report," *Organization for Economic Co-Operation and Development (OECD)*.

ALDASHEV, A., J. GERNANDT, AND S. L. THOMSEN (2012): "The Immigrant-Native Wage Gap in Germany," *Journal of Economics and Statistics (Jahrbuecher fuer Nationaloekonomie und Statistik), Justus-Liebig University Giessen, Department of Statistics and Economics*, 232(5), 490–517.

ALGAN, Y., C. DUSTMANN, A. GLITZ, AND A. MANNING (2010): "The Economic Situation of First and Second-Generation Immigrants in France, Germany and the United Kingdom," *Economic Journal, Royal Economic Society*, 120(542), F4–F30.

AZARIADIS, C., AND A. DRAZEN (1990): "Threshold Externalities in Economic Development," *The Quarterly Journal of Economics, MIT Press*, 105(2), 501–26.

BAKKER, D., M. ANDRE, V. VIVEKA, W. SOREN, B. CECIL H., B. PAMELA, D. EGOROV, M. ROBERT, G. ANTHONY, AND H. ERIC W. (2009): "Adding typology to lexicostatistics: A combined approach to language classification," *Lingustic Typology*, 12(1), 169–181.

BARTOLUCCI, C. (2009): "Understanding the Native-Immigrant Wage Gap Using Matched Employer-Employee Data. Evidence from Germany," .

BELL, B. D. (1997): "The Performance of Immigrants in the United Kingdom: Evidence from the GHS," *The Economic Journal*, 107(441), 333–344.

BILGER, F., R. JÄCKLE, B. VON ROSENBLADT, AND A. STRAUSS (2012): "Studiendesign, Durchführung und Methodik der leo. – Level-One Studie," in *Funktionaler Analphabetismus in Deutschland, Ergebnisse der ersten leo. – Level-One Studie*, ed. by A. Grotlüschen, and W. Riekmann, pp. 77–105. Waxmann Verlag.

BLEAKLEY, H., AND A. CHIN (2004): "Language Skills and Earnings: Evidence from Childhood Immigrants," *The Review of Economics and Statistics*, 86(2), 481–496.

BORJAS, G. J. (1985): "Assimilation, Changes in Cohort Quality, and the Earnings of Immigrants," *Journal of Labor Economics, University of Chicago Press*, 3(4), 463–89.

CARLINER, G. (1981): "Wage Differences by Language Group and the Market for Language Skills in Canada," *Journal of Human Resources*, 16(3), 384–399.

CHISWICK, B. R. (1978): "The Effect of Americanization on the Earnings of Foreign Born Men," *Journal of Political Economy*, 96(55), 897–922.

———— (1979): "The Economic Progress of Immigrants: Some Apparently Universal Patterns," in *Contemporary Economic Problems*, ed. by W. Fellner, pp. 357–399. American Enterprise Institute, Washington.

CHISWICK, B. R. (1980): "The Earnings of White and Coloured Male Immigrants in Britain," *Economica, London School of Economics and Political Science*, 47(185), 81–87.

CHISWICK, B. R. (1991): "Speaking, Reading, and Earnings among Low-skilled Immigrants," *Journal of Labor Economic*, 9(2), 149–170.

———— (1992): "Language in the Immigrant Labor Market:The Immigrant Experience in Canada and the United States," in *Immigration, Language and Ethnicity: Canada and the United States, Washington*, ed. by B. Chiswick. American Enterprise Institute.

CHISWICK, B. R., AND P. W. MILLER (1995): "The Endogeneity between Language and Earnings: International Analyses," *Journal of Labor Economics, University of Chicago Press*, 13(2), 246–88.

CHISWICK, B. R., AND P. W. MILLER (2005): "Linguistic Distance: A Quantitative Measure of the Distance Between English and Other Languages," *Journal of Multilingual and Multicultural Development*, 26(1), 1–11.

———— (2008): "A Test of the Critical Period Hypothesis for Language Learning," *Journal of Multilingual and Multicultural Development*, 29(1), 16–29.

———— (2012): "Negative and Positive Assimilation, Skill Transferability, and Lingustic Distance," *Journal of Human Capital*, 6(1), 35–55.

———— (2014): "International Migration and the Economics of Language," *IZA Discussion Papers*, 7880.

COULOMBE, S., AND J.-F. TREMBLAY (2006): "Literacy and Growth," *The B.E. Journal of Macroeconomics, De Gruyter*, 6(2), 1–34.

COULOMBE, S., J.-F. TREMBLAY, AND S. MARCHAND (2004): *Literacy scores, human capital and growth across fourteen OECD countries*. Statistics Canada Ottawa.

DENNY, K., C. HARMON, AND M. ROCHE (1997): "The distribution of discrimination in immigrant earnings-evidence from Britain 1974-1993," .

DUSTMANN, C. (1993): "Earnings Adjustment of Temporary Migrants," *Journal of Population Economics, Springer*, 6(2), 153–68.

DUSTMANN, C. (1994): "Speaking Fluency, Writing Fluency and Earnings of Migrants," *Journal of Population Economics*, 7, 133–156.

DUSTMANN, C., AND F. FABBRI (2003): "Language Proficiency and Labour Market Performance of Immigrants in the UK," *Economic Journal, Royal Economic Society*, 113(489), 695–717.

DUSTMANN, C., T. FRATTINI, AND I. P. PRESTON (2013): "The Effect of Immigration along the Distribution of Wages," *Review of Economic Studies*, 80(1), 145–173.

DUSTMANN, C., AND A. GLITZ (2011): "Migration and Education," in *Handbook of the Economics of Education*, ed. by E. A. Hanushek, M. Stephen, and W. Ludger, vol. 4, pp. 327–439. Elsevier B.V.

DUSTMANN, C., AND A. VAN SOEST (2001): "Language Fluency and Earnings: Estimation with Misclassied Language Indicators," *Review of Economics and Statistics*, 83(4), 663–674.

DUSTMANN, C., AND A. VAN SOEST (2002): "Language and the Earnings of Immigrants," *Industrial and Labor Relations Review*, pp. 473–492.

FERRER, A., D. A. GREEN, AND W. C. RIDDELL (2006): "The Effect of Literacy on Immigrant Earnings," *The Journal of Human Resources*, 41(2), 380–410.

FINNIE, R., AND R. MENG (2005): "Literacy and labour market outcomes: self-assessment versus test score measures," *Applied Economics, Taylor & Francis Journals*, 37(17), 1935–1951.

FRIEDBERG, R. M. (2001): "The impact of mass migration on the Israeli labor market," *Quarterly Journal of Economics*, pp. 1373–1408.

GREEN, D. A., AND C. RIDDELL (2003): "Literacy and earnings: an investigation of the interaction of cognitive and unobserved skills in earnings generation," *Labour Economics, Elsevier*, 10(2), 165–184.

GROTLÜSCHEN, A., AND W. RIEKMANN (2011): "leo.-Level–One Study, Literacy of adults at the lower rungs of the ladder.," *Press brochure, University Hamburg, http://blogs.epb.uni-hamburg.de/leo*.

HANUSHEK, E. A., AND L. WOESSMANN (2008): "The Role of Cognitive Skills in Economic Development," *Journal of Economic Literature*, 46(3), 607–68.

HARTIG, J., AND W. RIEKMANN (2012): "Bestimmung der Level-Grenzen in der leo. – Level-One Studie," in *Funktionaler Analphabetismus in Deutschland, Ergebnisse der ersten leo. – Level-One Studie*, ed. by A. Grotlüschen, and W. Riekmann, pp. 77–105. Waxmann Verlag.

ISPHORDING, I., AND S. OTTEN (2014): "Linguistic Barriers in the Destination Language Acquisition of Immigrants," *Journal of Economic Behavior and Organization*, 105, 30–50.

ISPHORDING, I. E. (2014): "Disadvantages of linguistic originEvidence from immigrant literacy scores," *Economics Letters*, 123(2), 236–239.

JÄCKLE, R., AND O. HIMMLER (2010): "Health and Wages: Panel Data Estimates Considering Selection and Endogenity," *Journal of Human Resources*, 45(2), 364–406.

JUNKER, B., L. S. SCHOFIELD, AND L. J. TAYLOR (2012): "The use of cognitive ability measures as explanatory variables in regression analysis," *IZA Journal of Labor Economics*, 1(4).

KAAS, L., AND C. MANGER (2012): "Ethnic Discrimination in Germany's Labour Market: A Field Experiment," *German Economic Review, Verein fr Socialpolitik*, 13(1), 1–20.

MCINTOSH, S., AND A. VIGNOLES (2001): "Measuring and Assessing the Impact of Basic Skills on Labour Market Outcomes," *Oxford Economic Papers*, 53(3), 453–81.

MINCER, J. (1958): "Investment in Human Capital and Personal Income Distribution," *Journal of Political Economy*, 66(4), 281–302.

——— (1974): *Schooling, Experience, and Earnings*. National Bureau of Economic Resaearch, New York, 3rd edn.

MIRANDA, A., AND Y. ZHU (2013): "English Deficiency and the Native-Immigrant Wage Gap," *Economics Letters*, 118(1), 38–41.

OZDEN, C., I. C. NEAGU, AND A. MATTOO (2005): *Brain Waste? Educated Immigrants in the U.S. Labor Market*. The World Bank.

PISCHKE, J.-S. (1992): "Assimilation and the earnings of guestworkers in Germany," *ZEW Discussion Papers*, 92-17.

RUBIN, D. B. (1987): *Multiple Imputations for Non-Response in Surveys*. Wiley, New York.

SANDERS, C., AND R. LESSEM (2013): "The Native-Immigrant Wage Gap in the United States," Discussion paper.

SPRIETSMA, M. (2013): "Discrimination in grading: experimental evidence from primary school teachers," *Empirical Economics, Springer*, 45(1), 523–538.

STEVENS, G. (2004): "Using Census Data to Test the Critical-Period Hypothesis for Second-Language Acquisition," *Psychological Science*, 15(3), 215–216.

VIGNOLES, A., A. DE COULON, AND O. MARCENARO-GUTIERREZ (2010): "The value of basic skills in the British labour market," *Oxford Economic Papers*, p. gpq012.

WILEY, E., E. BIALYSTOK, AND K. HAKUTA (2004): "New Approaches to Using Census Data to Test the Critical-Period Hypothesis for Second-Language Acquisition," *Psychological Science*, 16(3), 341–343.