Working Paper
No. 12
11/2017

Michael Christl,
Monika Köppl-Turyna,
Phillipp Gnan

Wage Differences Between Immigrants and Natives in Austria: The Role of Literacy Skills

Abstract

- This paper analyzes wage differences between natives and immigrants in Austria. First, we show that for both groups, literacy skills are an important determinant of the hourly wage. In the second step, we show that differences in proficiency with respect to literacy can explain more than three log points of the total wage gap of 9.7 log points between natives and immigrants.

- When adding literacy skills to the wage decomposition, the discriminatory part vanishes completely, suggesting that the wage difference between immigrants and natives in Austria can be to a large extent explained.

- Furthermore, we account for a possible sample selection bias. After controlling for literacy skills, the unexplained part of the gap becomes statistically insignificant. The importance of literacy skills in explaining wage differences between natives and immigrants is robust across several sensitivity tests.

Keywords

- Wage, decomposition, gap, immigrants, natives, Austria.

Ein Working Paper der Agenda Austria

Wage Differences Between Immigrants and Natives in Austria: The Role of Literacy Skills*

Michael Christl[a,*], Monika Köppl–Turyna[a,b], Phillipp Gnan[a]

[a] *Agenda Austria, Schottengasse 1/3, 1010 Vienna, Austria*
[b] *Vienna University of Economics and Business, Institute for Analytical Economics, Welthandelsplatz 1, 1020 Vienna, Austria*

Abstract

This paper analyzes wage differences between natives and immigrants in Austria. First, we show that for both groups, literacy skills are an important determinant of the hourly wage. In the second step, we show that differences in proficiency with respect to literacy can explain more than three log points of the total wage gap of 9.7 log points between natives and immigrants. When adding literacy skills to the wage decomposition, the discriminatory part vanishes completely, suggesting that the wage difference between immigrants and natives in Austria can be to a large extent explained. Furthermore, we account for a possible sample selection bias. After controlling for literacy skills, the unexplained part of the gap becomes statistically insignificant. The importance of literacy skills in explaining wage differences between natives and immigrants is robust across several sensitivity tests.

JEL Classification: J71, J15
Keywords: wage, decomposition, gap, immigrants, natives, Austria

*Corresponding author. Email: michael.christl@agenda-austria.at
*The authors are grateful to Gerhard Reitschuler for his helpful comments.

1. Introduction

The worldwide trend towards globalization is closely linked to the free movement of labor. In a global world, workers tend to move to other countries, sometimes for a better chance of finding work. This raises the question whether natives and immigrants are treated equally in the labor market. The literature identifies lower wages for immigrants due to lower returns to foreign education or lower returns to experience gained in other countries (see, Aldashev et al. 2012, Dell'Aringa et al. 2015). Studies in Austria have shown that a substantial part of the difference in wages between natives and immigrants can be explained by various characteristics such as occupation, education, experience or industry. Nevertheless, a significant part of the difference in wages remains unexplained in these studies, raising the question of whether this unexplained gap is due to discrimination or unobserved characteristics.

Recent studies show that the skills of a population are a "key ingredient in modern knowledge based societies" (see, e.g., Hanushek et al. 2015). Nevertheless, evidence for the returns to skills is rare, since the data on skills are scarce. We use the PIAAC data set of the Organization for Economic Co-operation and Development (OECD) for Austria. The data set is unique since it covers variables that assess several aspects of human capital unobserved in other studies, such as literacy skills, numeracy skills and problem solving skills (in a technology-rich environment). Taking a closer look at the skill endowment of the Austrian population, the data set reveals substantial average differences between the skill endowments of natives and immigrants, raising the question of whether these differences can explain wage disparities between the two groups.

This paper makes a first attempt to analyze the impact of skills endowment on wage differences between natives and immigrants in Austria. In particular, we will focus on the role of literacy skills, since of the three classes of skills assessed in the PIAAC data set literacy skills naturally appear to have a predominant role in affecting wages, given their universal importance for a broad range of tasks. First, we analyze the impact of literacy skills on wages of both natives and immigrants. Thereafter, we use standard decomposition methods to separate the wage gap into explained and unexplained parts. We also control for sample selection, since immigrant women tend to participate less in the labor market. As a robustness check, we also employ a matching approach to overcome the sample selection issue and to account for differences in the support of the two groups with respect to their characteristics endowment.

We show that literacy skills substantially influence the wages of both natives and immigrants. Additionally, our analysis suggests that considering literacy skills can considerably narrow the unexplained part of the wage gap between these two groups in Austria, explaining between 20% and one third of the total wage gap between natives and immigrants.

This work is structured as follows. The next section gives an overview of the recent literature on wage gaps between native and migrant populations. Section 3 presents the empirical methods used together with details of the data set. Section 4 presents the results of the investigation. Section 5 concludes. Additional tables and figures are presented in the Appendix.

2. Previous Literature

Wage gaps are frequently discussed in the literature, mainly because economists are interested in unexplained differences between social groups. These unexplained differences may be due to either missing data at an individual level or discrimination.

Accordingly, the question of the existence and extent of wage discrimination against immigrants has naturally been raised in a large number of countries around the globe, leading to manifold studies and the use of various techniques to assess such a wage gap. A wide variety of studies have shown that wages and returns to human capital are lower for immigrants than for native-born workers (see, e.g., Chiswick 1978, Dustmann 1993, Chiswick and Miller 2008).

While the method of choice is often strongly influenced by data availability, the most common approaches are wage gap decompositions similar to those described by Oaxaca (1973) and Blinder (1973), where the basic idea is to explain a part of the gap by the individuals' differing characteristics in the two groups, leading to varying levels of productivity which should be rewarded accordingly by the employer. The "unexplained" part of the gap is then interpreted as discrimination against one of the groups, on the assumption that the assessed characteristics properly reflect an individual's productivity.

Recent research, e.g., that of Lehmer and Ludsteck (2011), employs this type of decomposition as well as quantile regression methods, further differentiating immigrants in Germany by origin from a range of countries inside and outside Europe. In a more recent report, Lehmer and Ludsteck (2013) stress the importance of the country of origin for the extent of wage assimilation. Again for the case of Germany, Aldashev et al. (2012) chose the Oaxaca-Blinder decomposition technique to show that merely distinguishing between foreigners and native Germans would result in assigning naturalized immigrants to the wrong group for this particular purpose. They find that a considerable wage gap exists for immigrants who received their education abroad, while being educated in Germany substantially decreases the immigrants' wage disadvantage. Similarly, for the Italian labor market, results from Dell'Aringa et al. (2015) show that pre-immigration work experience does not yield statistically significant returns (OLS regression with interaction dummies). Along similar lines, Coulombe et al. (2014) create a proxy for the quality of human capital (GDP per capita of the country of origin), which they include in an Oaxaca-Blinder-style analysis of the wage differential in Canada.

In contrast to the Oaxaca-Blinder approach, Bartolucci (2014) argues that the use of individuals' characteristics in the residual approach represents an insufficient proxy for productivity, and this author proposes to apply an approach extending the method of Hellerstein and Neumark (1999), where estimated productivity differentials between two groups are directly compared to their wage differentials. Bartolucci uses matched employer-employee data from Germany and concludes that immigrants are being discriminated against. However, he also finds that the wage gap reverses for immigrants in highly qualified positions.

Another decomposition approach is based on the matching procedure of Ñopo (2008). Garcia et al. (2009) employ this method to analyze racial wage gaps in Brazil, which are shown to be mainly due to differences in the observed characteristics. Nicodemo and Ramos (2012) use the same technique to examine wage discrimination against immigrant women in Spain. They find evidence for a wage gap that is mainly caused by immigrants' segregation into different occupations. Examining the long-term perspective for the same country, Izquierdo et al. (2009) conclude that the initial wage differential decreases steadily over the first five to six years of residence, but never disappears completely.

Another example from a long list of countries for which immigrant wage gap-related research has been conducted, is that of Australia, where Cai and Liu (2015) employ the semi-parametric method of DiNardo et al. (1995) and unconditional quantile regression to examine immigrants' wage differentials across the entire wage distribution. They find immigrants from English-speaking countries to be better off than native Australians, but also find evidence of disadvantages for non-English-speaking female immigrants.

While for many countries wage differences between immigrants and natives have been analyzed in depth, the literature on wage differences in Austria is scarce, despite the fact that people with an immigrant background in Austria accounted for 22.1% of the overall population in 2016[2].

This scarcity of research in Austria is mostly a result of data limitations, although some evidence indicates that immigrants face certain disadvantages in the labor market. Grandner and Gstach (2015) compare the wages of immigrants and natives in Austria using EU-SILC data. They employ the Oaxaca-Blinder approach, as well as the quantile regression approach described by Machado and Mata (2005), to disentangle the unexplained and explained parts of the wage gap between immigrants and natives.

According to Grandner and Gstach (2015), immigrants earn between 15% and 25% less than natives, where the highest difference in wages appears at the eighth decile of the wage distribution. Their results suggest that most of the wage gap in the lower end of the wage distribution can be explained. However, in the middle part of the wage distribution, the unexplained part reaches 12% and increases to 20% at the eighth decile.

Hofer et al. (2017) use data from the Austrian micro-census (Labour Force Survey). They follow the Oaxaca-Blinder approach and also use the counterfactual distribution approach of Chernozhukov et al. (2013) to extend the decomposition to the full wage distribution. They also distinguish between the wage gaps of immigrants and natives by gender. The wage differential lies between 7% (first decile) and 21% (ninth decile) for males, and between 5% (first decile) and almost 20% (eighth decile) for females. They show that the gap is reduced

[2] Austrian Statistical Office 2016.

to between zero and 8%, depending on the wage decile, for both males and females after controlling for certain characteristics.

Both papers indicate substantial differences in the total wage gap, but also in the explained part. The difference might be explained by the fact that Hofer et al. (2017) control for more individual and firm characteristics. Due to the lower number of observations, we do not extend the decomposition to the full wage distribution, but use Oaxaca-Blinder decomposition to split up the wage gap at an average level. To compare our results to other findings, we summarize the OB results of both papers in Table 1. While Grandner and Gstach (2015) found an unexplained wage gap of 9% at the average level, Hofer et al. (2017) are left with an unexplained gap of 2.9% for males and 5.0% percent for females.

Table 1: Wage gaps in Austria –Oaxaca-Blinder

	Grandner and Gstach (2015)	**Hofer et al. (2017)** male	**Hofer et al. (2017)** female
Total	-0.211	-0.136	-0.172
Explained	-0.121	-0.107	-0.122
Unexplained	-0.090	-0.029	-0.050
Immigrant definition	1st generation	1st and 2nd generation	1st and 2nd generation

3. Methodology and Data

3.1. Data

In our analysis, we use data from the PIAAC survey of the OECD. The data were collected in 2011/12 and they include 4,810 individual observations. After filtering out missing data, our sample reduces to about 2,500 observations[3].

The migration background variable of the PIAAC data set uses information on the country of birth of the respondents' parents, and allows us to distinguish between first- and second-generation immigrants and natives. In the further course of this paper, we will only consider first-generation immigrants and natives, since the number of observations of second-generation

[3]In terms of the number of observations, this is smaller than the EU-SILC data set used, for instance, by Grandner and Gstach (2015), or the Austrian micro-census data set (matched with social security records) used by Hofer et al. (2017).

immigrants in the sample is rather small, and also, they are already much more assimilated compared to first-generation immigrants, making a clear-cut distinction more problematic.

The dependent variable in all estimations is the hourly wage[4]. As in previous studies, detailed information about personal characteristics (such as age, job experience, education, children, family status, etc.) and job characteristics (such as occupation, the size of the firm, hours worked, etc.) are provided. The greatest advantage of the PIAAC data set, however, is the possibility of controlling for various skills, which is typically not possible using other data sets. Additionally, we can control for specific task profiles, which, according to Autor and Handel (2013), differ substantially even within narrowly defined occupational groups.

Unlike, for instance, Hofer et al. (2017), who use job characteristics such as unskilled, low-skilled, medium-skilled etc., we can directly control for the skill proficiency and skill use of immigrants and natives which directly affect productivity and should therefore be rewarded by the employer. More precisely, the PIAAC survey assesses proficiency in literacy, numeracy and problem-solving skills[5]. Literacy proficiency, being closely tied to seamless and efficient communication in professional life and being universally applicable to, and required for, a broad range of situations and tasks, certainly proves to be especially crucial for successful integration into the Austrian labor market and into society in general (see, e.g., Derwing and Waugh 2012, Dustmann 1993). Therefore, this paper will focus especially on the role of literacy skills, leaving a closer examination of the other two classes of skills for further research. In the PIAAC data set, literacy captures the respondents' ability to comprehend and interact with text.

Table 2 gives a summary of the statistics of our sample. The average hourly gross wage is about 13.04 euro for immigrants, while natives earn approximately 14.32 euro on average. Additionally, we can see substantial differences between immigrants and natives in many personal and job-related characteristics.

Regarding the educational level, on the one hand, we can see that 26% of the immigrants

[4]To see whether the wage distribution of the PIAAC data set is representative of Austrian wage distribution, we compare the wage data of PIAAC with the wage tax statistics from Statistics Austria (see Figure 4 in the Appendix).

[5]For a detailed description of the methodology and assessed aspects of these skills, see OECD (2016).

Table 2: Summary statistics

	natives			immigrants		
Variable	Observations	Mean	Std. Dev.	Obs	Mean	Std. Dev.
Hourly wage	2246	14.24	8.31	366	12.94	8.07
Male	3723	0.50	0.50	620	0.49	0.50
Children	3723	0.60	0.49	619	0.70	0.46
Experience	3501	19.83	12.77	584	16.83	11.40
Hours worked	2815	37.94	12.83	431	37.46	12.09
Education						
ISCED1	3723	0.01	0.10	620	0.05	0.22
ISCED2	3,723	0.16	0.37	620	0.21	0.41
ISCED3	3,723	0.49	0.50	620	0.38	0.49
ISCED4	3,723	0.14	0.34	620	0.10	0.30
ISCED56	3,723	0.20	0.40	620	0.26	0.44
Sector						
Private sector	2808	0.73	0.44	431	0.81	0.39
Public sector	2808	0.24	0.43	431	0.16	0.37
NGO	2808	0.03	0.17	431	0.03	0.17
Firm size						
Up to 10	2431	0.26	0.44	386	0.26	0.44
11 to 50	2431	0.31	0.46	386	0.31	0.46
51 to 250	2431	0.21	0.41	386	0.24	0.43
251 to 1,000	2431	0.14	0.35	386	0.12	0.33
More than 1,000	2431	0.07	0.26	386	0.08	0.26
Occupation						
Armed forces	2,766	0.01	0.08	421	0.00	0.00
Managers	2,766	0.06	0.24	421	0.07	0.25
Professionals	2,766	0.18	0.38	421	0.17	0.38
Technicians	2,766	0.22	0.42	421	0.15	0.36
Clerks	2,766	0.11	0.31	421	0.06	0.24
Service workers	2,766	0.15	0.36	421	0.17	0.38
Skilled agricultural workers	2,766	0.05	0.21	421	0.02	0.14
Craft workers	2,766	0.12	0.33	421	0.10	0.31
Machine operators	2,766	0.05	0.22	421	0.09	0.28
Elementary occ.	2,766	0.05	0.21	421	0.16	0.37
Literacy skills	3,723	2.77	0.38	620	2.50	0.53
Work tasks						
Writing tasks	2,984	2.69	1.65	488	2.20	1.80
Reading tasks	2,984	3.02	1.52	488	2.44	1.72
Numerical tasks	2,984	2.42	1.73	488	1.94	1.84

have no more than primary education (ISCED1), while only 17% of the natives are in this category. On the other hand, the share of workers with tertiary education (ISCED56) is higher for immigrants (26%) compared to natives (20%). The proportion of men is similar for immigrants and natives, although it should be noted that immigrant women tend to exhibit a lower labor-market participation rate than immigrant men (this is true of all observations, regardless of working status). Furthermore, 70% of immigrants have children, compared to only 60% of the native population.

On average, immigrants have three years less work experience compared to natives, which is likely to be a result of higher unemployment compared to natives.

Concerning job-related characteristics, we see almost no difference in firm size between immigrant and native employees. However, there are significant disparities in the occupations and sectors immigrants and natives work in. While a large share of immigrants work as machine operators or in elementary occupations, a higher share of natives work as professionals, technicians or clerks. In contrast, a slightly higher share of immigrants work in managerial occupations compared to natives.

Additionally, most immigrants work in the private sector (81%). There is a significant difference in public sector employment between natives and immigrants. While 24% of natives work in the public sector, only 16% of immigrants do so.

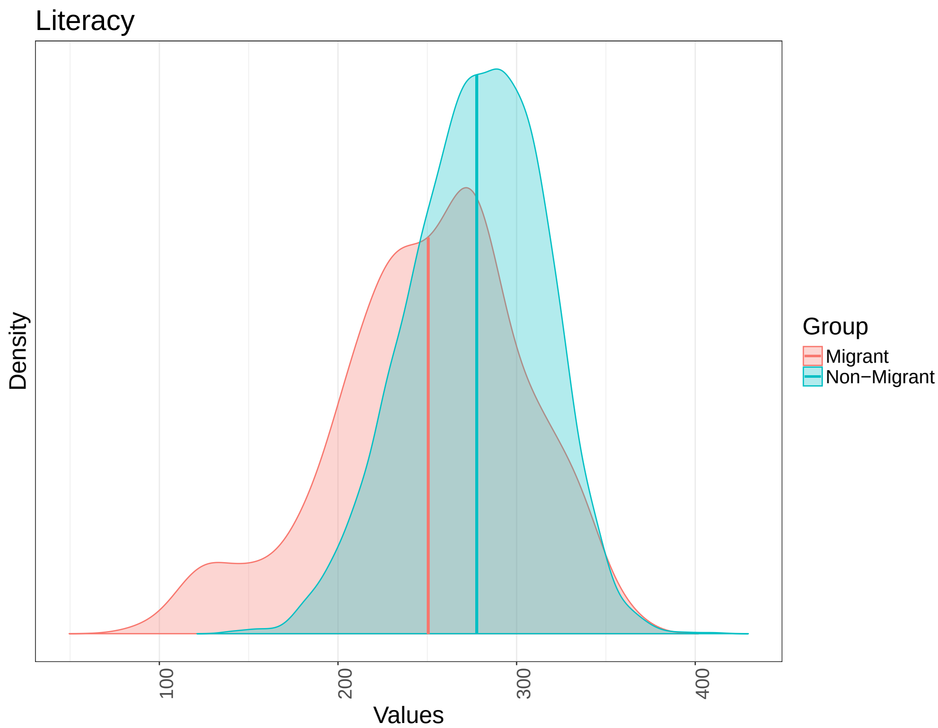
Concerning work tasks, we can see that natives are required to use reading, writing and numerical skills more often than immigrants. The PIAAC questionnaire assesses the frequency of these tasks at work. The respondents answers are first used to calculate indices, and subsequently partitioned into quintiles.

Turning to literacy, our characteristic of particular interest, we can see substantial differences between natives and immigrants. The difference in mean literacy scores between the two groups (2.77 for natives and 2.50 for immigrants) is statistically highly significant (p-value $<$ 2.2e-16). Figure 1 shows the corresponding empirical distributions. For immigrants, we find that not only is the center of the distribution further to the left, but also that the left tail of the distribution is much heavier, meaning that many immigrants have only rudimentary literacy skills.

3.2. Methodology

The goal of our paper is to examine the wage differences between natives and immigrants in Austria, with particular regard to the role of literacy skills. There are several methods for decomposing the wage gap into an explained part and an unexplained part. A selection of approaches is summarized in Fortin et al. (2011). Due to the restricted sample size of the PIAAC data set, we use the Oaxaca-Blinder decomposition to decompose the mean wage differences across two groups. As a robustness check, we also use a new matching technique

Figure 1: Empirical distribution of literacy skills: differences between natives and immigrants in Austria



introduced by Ñopo (2008) that relies on fewer assumptions than the OB decomposition, but is limited by the sample size.

3.2.1. Oaxaca-Blinder Decomposition

The OB method decomposes the wage gap between immigrants and natives into two components: the gap that results from differences in productivity-related characteristics and a residual, often called the "unexplained residual". Mincer wage regressions are estimated for both groups separately ($Y_i = \beta_i X_i + \varepsilon_i$), where the coefficients typically represent the return (price) of certain productivity-related characteristics. The difference in the means can therefore be written as:

$$\ln \overline{Y}_M - \ln \overline{Y}_N = \overline{X}_M \beta_M - \overline{X}_N \beta_N \tag{1}$$

where $\ln \overline{Y}_i$ denotes the average hourly log-wage for immigrants ($i = M$) and natives ($i = N$), $\overline{X}_i$ the corresponding vector of the mean values of the characteristics and β_i the vector of coefficients. Assuming that the wage structure of the natives is the non-discriminatory wage distribution[6], we obtain:

$$\ln \overline{Y}_M - \ln \overline{Y}_N = \left(\overline{X}_M - \overline{X}_N\right) \beta_N + \overline{X}_M \left(\beta_M - \beta_N\right) \tag{2}$$

where the first part is the wage gap due to differences in mean characteristics $(\overline{X})$ and the second part corresponds to differences in returns, often interpreted as discrimination since the same characteristics are rewarded differently in the two groups.

3.2.2. Matching

As noticed by Ñopo (2008), the standard Blinder (1973) and Oaxaca (1973) decomposition ignores the fact that the supports of the empirical distributions of individual characteristics between the two groups might be different (see also Heckman et al. 1997). By not considering the differences in the supports, the Oaxaca-Blinder approach requires the "out-of-support assumption", i.e., that the estimates are also valid outside the common support. Whenever systematic differences between the out-of- and in-support observations persist, the standard decomposition empirically tends to overestimate the component of the gap attributable to coefficients.

Ñopo (2008) instead proposes a matching algorithm, which requires neither assumptions about the support nor estimations of the earnings equations. This decomposition further allows the differences in the distributions of the individuals' characteristics to be accounted for. The procedure works as follows.

1. Select one treatment observation (immigrant) from the sample (without replacement).
2. Select all control-group observations (natives) with the same characteristics.
3. Using the selected natives, construct a synthetic native, whose wage is the average of all the selected natives and match the synthetic native to the original immigrant.
4. Repeat until the treatment group has been exhausted.

[6]In our calculations we use the two-fold decomposition that uses the pooled model excluding group variance.

As a result, the data set is partitioned into four groups: matched immigrants, matched natives, unmatched immigrants and unmatched natives. The overall gap Δ can be now decomposed according to

$$\Delta = \Delta_M + \Delta_X + \Delta_0 + \Delta_F, \tag{3}$$

where

- Δ_M corresponds to the difference between the unmatched and the matched natives,
- Δ_X corresponds to differences in characteristics between matched natives and immigrants,
- Δ_0 corresponds to differences in returns in the matched group, i.e., the "unexplained part",
- Δ_F is the difference between the characteristics of the matched and the unmatched immigrants.

In other words, the sum $\Delta_M + \Delta_X + \Delta_F$ corresponds to the part of the gap which is due to differences in characteristics, similar to the traditional OB decomposition, but allows the differences between matched and unmatched individuals to be inferred. Note, however, that only Δ_X can be interpreted as the "explained" part of the raw gap (so to speak, *"justified"*) by differences in characteristics.

4. Empirical Findings

4.1. Wage Regressions

To analyze the immigrant wage gap in Austria, we first estimate three different specifications of a linear regression model which we will subsequently employ in the Oaxaca-Blinder decomposition. Similar to Hofer et al. (2017), we use all individual characteristics (experience, education,...) as well as the firm- and job-related characteristics (firm size, sector,...) of our data set as explanatory variables in the first specification[7]. In the second specification, we

[7]Note that Hofer et al. (2017) use more characteristics in their data set.

add the literacy proficiency[8] and in the third specification, we also control for work tasks. The second and third specifications should therefore not only control for typical individual and firm-specific characteristics, but also for human capital (literacy) not so far controlled for in other studies. For a detailed description of the variables employed to capture an individual's characteristics, see Table 15 in the Appendix.

Table 3 summarizes the estimated coefficients from log-linear regression models with robust standard errors for all three specifications (with log of hourly wage as the dependent variable). Columns (1) to (3) correspond to natives, columns (4) to (6) to immigrants and column (7) shows the results for the most comprehensive model specification using the pooled sample of immigrants and natives together.

While the difference in wages between men and women can be interpreted to be of a similar nature for natives and immigrants, amounting to an increase for men compared to women of roughly 14 and 13 log points for natives and immigrants respectively (specification 3), the return on work experience seems to be different for the two groups. The corresponding linear term is estimated to be roughly 0.045 for natives, but only around 0.013 for immigrants (even though this is partly offset by a less negative coefficient of the quadratic term for immigrants), pointing in the same direction as previous findings by Dell'Aringa et al. (2015) for the Italian labor market. The coefficients for education by ISCED classification relative to the ISCED1 level (primary education or below) suggest that a high level of education yields higher returns for natives compared to immigrants, although this advantage considerably decreases when controlling for tasks.

By contrast, when considering the ISCO1C variable accounting for occupation, immigrants from levels 1 to 4 (broadly speaking, white-collar occupations) tend to experience higher wage increases over the baseline (elementary occupations) than do natives. However, this might partly be attributed to a more pronounced pay gap in lower-qualification occupations (by about seven log points), i.e., immigrants belonging to the category "elementary occupations" start out at a lower level than their native counterparts.

[8] Adding problem-solving skills has no statistical significance and reduces the sample size further, therefore we do not include it in the decomposition analysis. Results are available upon request.

Turning to the sectoral aspect of employment, the coefficients hint at no significant differences for immigrants between the private, public and NGO sectors, but significantly lower wages in the public sector for natives compared to the private sector (baseline). It is also noteworthy that only 18% of immigrants in the sample are employed in the public sector, compared to 27.3% of natives. Concerning firm size, natives gain strongly and steadily from increased firm size. The effect for immigrants is also positive, but much more modest and mostly not statistically significant.

Examining our work tasks variables, we can see that the effects of the frequency of writing, reading and numerical tasks are all positive and significant for natives. By contrast, none of these tasks variables is significant for immigrants. Finally, with respect to literacy, at a first glance the regressions similarly suggest that higher literacy skills only yield a positive return for natives (around 15 log points on wages per increase of 100 index points), and that there is no significant compensation for immigrants at all. However, this rather unintuitive result might hint at the different kinds of occupations into which immigrants and natives are generally segregated. In the next step, we therefore allow for interaction between the variable capturing occupation and the literacy index variable, to obtain a more detailed picture of the effect of literacy on wages.

4.1.1. Returns to Literacy Skills in Certain Occupations

Given that the absence of a positive effect of literacy skills on immigrants' wages in the above regressions might arise due to the fact that immigrants are overrepresented in occupations with no necessity for literacy skills, we extend the most comprehensive model specification from the above (specification 3) with an interaction term between occupation and literacy. Significant positive coefficients of the interaction term can be interpreted as denoting an occupation in which literacy skills are rewarded by the market.

The coefficients of other variables are very similar to the results shown above. Table 4 concentrates on the coefficients for occupation and literacy as well as their interaction[9]. All interaction terms are also visualized in Figure 2. We can see that for natives, managers earn

[9]Results for all coefficients are available upon request.

Table 3: Wage regression results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Male	0.1760***	0.1600***	0.1434***	0.1374***	0.1421***	0.1316***	0.1415***
	(9.42)	(8.60)	(7.76)	(3.22)	(3.33)	(3.08)	(8.31)
Children	-0.0015	-0.0029	-0.0087	0.0571	0.0697	0.0635	0.0067
	(-0.07)	(-0.15)	(-0.45)	(1.26)	(1.51)	(1.39)	(0.37)
Hours worked	-0.0050***	-0.0050***	-0.0066***	-0.0077***	-0.0077***	-0.0087***	-0.0070***
	(-6.27)	(-6.40)	(-8.31)	(-4.19)	(-4.18)	(-4.67)	(-9.55)
Experience	0.0448***	0.0451***	0.0438***	0.0137**	0.0136**	0.0126*	0.0390***
	(16.37)	(16.64)	(16.37)	(2.01)	(2.01)	(1.87)	(15.63)
Experience^2	-0.0007***	-0.0007***	-0.0006***	-0.0002	-0.0002	-0.0001	-0.0006***
	(-11.06)	(-10.68)	(-10.38)	(-1.00)	(-1.02)	(-0.84)	(-9.83)
ISCED1				baseline			
ISCED2	-0.1929*	-0.2031*	-0.2158**	-0.1239	-0.1376	-0.1338	-0.2047***
	(-1.74)	(-1.86)	(-2.00)	(-1.23)	(-1.37)	(-1.34)	(-2.82)
ISCED3	0.0694	0.0368	0.0056	0.0428	0.0096	0.0017	-0.0027
	(0.64)	(0.34)	(0.05)	(0.44)	(0.10)	(0.02)	(-0.04)
ISCED4	0.2342**	0.1719	0.1228	0.1241	0.0820	0.0860	0.1047
	(2.11)	(1.56)	(1.13)	(1.15)	(0.74)	(0.78)	(1.41)
ISCED56	0.3406***	0.2709**	0.2121*	0.2000*	0.1517	0.1442	0.1946***
	(3.06)	(2.45)	(1.94)	(1.90)	(1.39)	(1.32)	(2.60)
Armed forces	-0.3153***	-0.3567***	-0.4420***	-	-	-	-0.4040***
	(-3.11)	(-3.56)	(-4.44)	-	-	-	(-4.12)
Managers	0.3410***	0.2868***	0.1655***	0.7499***	0.7088***	0.6103***	0.2400***
	(6.69)	(5.63)	(3.17)	(7.88)	(7.21)	(5.79)	(5.18)
Professionals	0.2705***	0.2216***	0.1132**	0.4757***	0.4412***	0.3652***	0.1603***
	(6.02)	(4.93)	(2.45)	(6.01)	(5.39)	(4.18)	(3.97)
Technicians	0.1560***	0.1175***	0.0152	0.4072***	0.3852***	0.3022***	0.0637*
	(3.82)	(2.88)	(0.36)	(5.56)	(5.19)	(3.77)	(1.75)
Clerks	0.1106***	0.0719*	-0.0322	0.3035***	0.2739***	0.2136**	0.0070
	(2.60)	(1.69)	(-0.74)	(3.46)	(3.06)	(2.33)	(0.18)
Service workers	-0.0456	-0.0637	-0.1170***	0.0449	0.0307	0.0068	-0.0903***
	(-1.12)	(-1.58)	(-2.90)	(0.68)	(0.46)	(0.10)	(-2.63)
Skilled agricultural workers	-0.0648	-0.0722	-0.0854	-0.0292	-0.0345	-0.0557	-0.0770
	(-0.71)	(-0.80)	(-0.95)	(-0.23)	(-0.27)	(-0.43)	(-1.03)
Craft workers	-0.0406	-0.0466	-0.0821*	0.0937	0.0966	0.0772	-0.0525
	(-0.95)	(-1.10)	(-1.95)	(1.28)	(1.32)	(1.06)	(-1.46)
Machine operators	-0.0109	-0.0113	-0.0110	0.0264	0.0202	0.0141	-0.0126
	(-0.22)	(-0.23)	(-0.23)	(0.34)	(0.26)	(0.18)	(-0.31)
Elementary occ.				baseline			
Up to 10				baseline			
11 to 50	0.0992***	0.0926***	0.0964***	0.0095	0.0112	0.0207	0.0857***
	(4.70)	(4.44)	(4.69)	(0.19)	(0.23)	(0.42)	(4.48)
51 to 250	0.1451***	0.1343***	0.1351***	0.0513	0.0433	0.0557	0.1242***
	(6.10)	(5.70)	(5.81)	(0.94)	(0.80)	(1.01)	(5.76)
251 to 1,000	0.1949***	0.1767***	0.1847***	0.1960***	0.1808***	0.1865***	0.1862***
	(7.30)	(6.67)	(7.06)	(2.96)	(2.71)	(2.77)	(7.58)
More than 1,000	0.2379***	0.2271***	0.2288***	0.0866	0.0810	0.0739	0.2144***
	(7.19)	(6.93)	(7.09)	(1.08)	(1.01)	(0.93)	(7.12)
Literacy		0.0018***	0.0016***		0.0008	0.0004	0.0012***
		(7.11)	(6.16)		(1.62)	(0.84)	(5.49)
Writing tasks			0.0145**			0.0204	0.0150**
			(2.22)			(1.26)	(2.45)
Reading tasks			0.0333***			0.0094	0.0294***
			(4.29)			(0.51)	(4.08)
Numerical tasks			0.0135**			0.0147	0.0154***
			(2.34)			(0.99)	(2.85)
Immigrant							-0.0055
							(-0.25)
Constant	1.8329***	1.3869***	1.4784***	2.1234***	1.9683***	2.0585***	1.6165***
	(16.15)	(10.79)	(11.63)	(18.21)	(13.08)	(13.44)	(17.61)
Observations	2,133	2,133	2,133	351	351	351	2,484
add. control variables not reported:	*sector*						

Robust t-statistics in parentheses; significance *0.1 **0.05 ***0.01

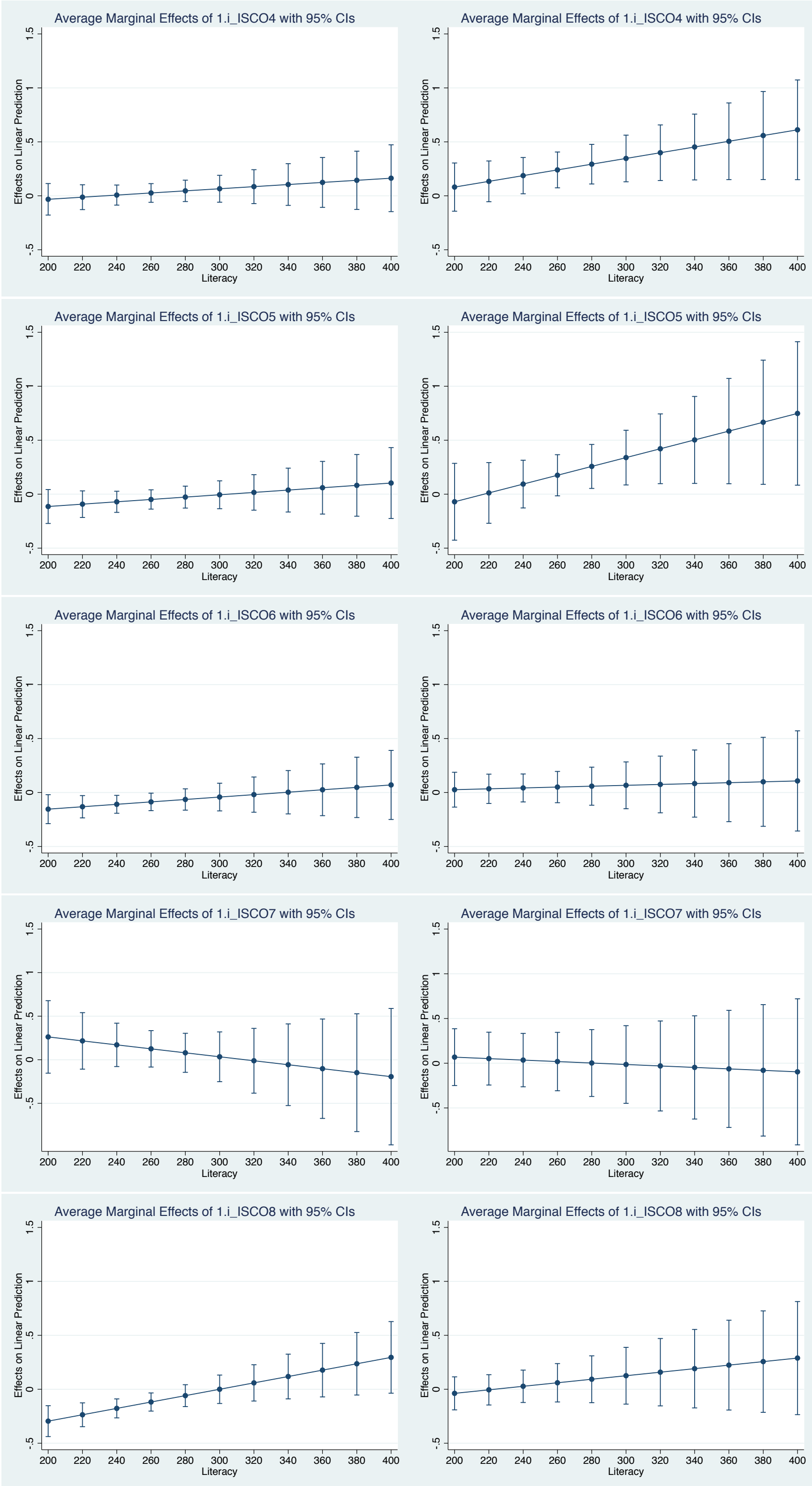
significant returns on higher literacy skills, while the overall effect of literacy (unconditional literacy plus literacy x occupation) is positive for most of the groups (except for ISCO0 and ISCO6), hinting at some general positive effect of literacy skills on natives' wages. For immigrants, the picture appears to be more mixed, with a positive relation between literacy and wages only observed for white-collar occupations (ISCO1-ISCO4, but excluding ISCO2), while there seems to be no additional payoff for literacy skills in jobs that typically require less formal education (ISCO5-9). However, as the coefficients corresponding to the interaction terms in ISCO1, ISCO3 and ISCO4 are statistically significant, this clearly shows that in certain groups of occupations, literacy does indeed play an important role for immigrants.

The above findings indicate that literacy skills play a crucial role for natives in general, but also for immigrants in certain occupations. In most typical white-collar occupations, literacy skills increase the wages of immigrants. Furthermore, certain levels of literacy skills might be demanded to obtain such jobs. As a consequence, not only the difference in literacy skills, but also the returns to literacy skills could play an important role in explaining the wage differential. Therefore, it seems reasonable to take them into account when analyzing wage differences between immigrants and natives in the next step.

Table 4: Wage regression with literacy-occupation interaction

	Natives	Immigrants	Both
Armed forces	1.6377*	-	1.4693*
	(1.87)	-	(1.71)
Managers	-1.2661***	-0.5868	-1.3103***
	(-3.14)	(-0.92)	(-4.03)
Professionals	0.2266	0.8419**	0.1676
	(0.74)	(2.07)	(0.74)
Technicians	-0.0593	-0.4667	-0.3078
	(-0.20)	(-1.18)	(-1.47)
Clerks	-0.1687	-1.4252**	-0.4633*
	(-0.53)	(-2.14)	(-1.89)
Service workers	-0.1690	-0.1068	-0.2923
	(-0.58)	(-0.31)	(-1.43)
Skilled agricultural workers	0.8592	-0.0654	0.3900
	(1.10)	(-0.12)	(0.89)
Craft workers	-0.5968*	-0.2719	-0.6396***
	(-1.95)	(-0.81)	(-3.05)
Machine operators	0.1177	0.4210	-0.1881
	(0.36)	(0.88)	(-0.78)
Elementary occupations		baseline	
Literacy	0.0011	-0.0004	0.0000
	(1.09)	(-0.43)	(0.02)
Armed forces × Literacy	-0.0070**		-0.0061**
	(-2.32)		(-2.09)
Managers × Literacy	0.0048***	0.0043*	0.0054***
	(3.36)	(1.93)	(4.75)
Professionals × Literacy	-0.0003	-0.0013	0.0002
	(-0.24)	(-0.89)	(0.29)
Technicians × Literacy	0.0003	0.0030**	0.0015*
	(0.29)	(1.97)	(1.86)
Clerks × Literacy	0.0005	0.0062**	0.0019**
	(0.45)	(2.47)	(2.00)
Service workers × Literacy	0.0002	0.0006	0.0009
	(0.19)	(0.40)	(1.07)
Skilled agricultural workers × Literacy	-0.0035	0.0001	-0.0018
	(-1.20)	(0.02)	(-1.03)
Craft workers × Literacy	0.0020*	0.0016	0.0023***
	(1.65)	(1.09)	(2.82)
Machine operators × Literacy	-0.0005	-0.0017	0.0008
	(-0.38)	(-0.83)	(0.78)
Constant	1.5859***	2.2095***	1.9162***
	(5.72)	(9.45)	(11.23)
Observations	2,133	351	2,484

Figure 2: Marginal effects of Literacy in different occupations (natives left, immigrants right)



4.2. The Immigrant Wage Gap

Using the variable specifications from the wage regressions above (Tables 3 and 4), we now employ the Oaxaca-Blinder method to further examine the wage differential and disentangle the effects of various characteristics on the wage gap[10].

Table 5 summarizes the decomposition of the wage gap into explained and unexplained parts. The variation by specification of roughly between 5.1 (specification 1) and 9.3 (specification 3) log points out of the total gap of 9.7 log points can be attributed to differences in the individuals' characteristics, while around 4.6 to 0.5 log points remain unexplained. Note that our unexplained gap is shown to be smaller in specifications 2 and 3 compared to the results of Hofer et al. (2017), but that their raw wage gap is also higher at 15.2 and 18.1 log points for men and women respectively, compared to just 9.7 log points for our sample. The main reason for this discrepancy is that while their study only considers individuals aged 20 to 55, our sample includes people aged 16 to 65. Section 4.4.1 will discuss results for a restricted age group (age $\geq$ 20), where a more similar raw gap is observed. Overall, our results show that although in specification 1 the unexplained part of the wage gap is still statistically significant at the 0.05 level, controlling for literacy skills yields an unexplained part that is no longer significantly different from zero, i.e., the unexplained part of the wage gap closes. Adding the tasks variables further reduces the unexplained wage differential, while the results do not change meaningfully when extending specification 3 by the interaction term (literacy x occupation).

Table 5: Oaxaca-Blinder decomposition

	Specification 1		Specification 2		Specification 3		Interaction	
Difference	0.0973***	(3.56)	0.0973***	(3.56)	0.0973***	(3.56)	0.0973***	(3.56)
Explained	0.0512**	(2.27)	0.0846***	(3.58)	0.0925***	(3.88)	0,0904***	(3.83)
Unexplained	0.0461**	(2.23)	0.0127	(0.62)	0.0048	(0.24)	0.0069	(0.35)

The contributions of the individual variables to the overall gap are listed in Table 6. The results for the specifications including the interaction terms are not discussed separately, as

[10] In particular, the results presented below are based on coefficients from a pooled model (Stata's omega implementation without the *group* control variable).

the overall explained gap remains almost identical to specification 3 and the contributions of literacy and occupation are no longer clearly separable (see Table 16 in the Appendix). Clearly, literacy skills appear to be a valuable indicator of human capital, strongly affecting the hourly wage. Accordingly, differences in literacy proficiency between the two groups can explain more than a third of the total wage gap. Considering specification 3 (1), disparities in work experience account for another 27.3% (25.7%) of the total gap, and disparities in education for 7.2% (12.4%). Furthermore, differences in the distribution across occupations and the frequency of various work tasks can together explain 41% (32.5%) of the overall wage gap. However, it should be noted that the selection of employees in different sectors and occupations could provide room for discrimination in itself. Differentiating between wage discrimination (where comparable immigrants and natives in a similar position are paid differently) and job discrimination (where comparable immigrants and natives are not offered the same jobs), this analysis focuses purely on the former, while also acknowledging the need for further examination of the latter.

The negative sign of the part of the gap explained by gender can be attributed to the different ratios of men and women in the two groups: even though the ratios are almost identical in the overall PIAAC data set, immigrant women exhibit a much lower labor-force participation rate (60%) compared to native women (74%). Men are paid more by roughly the same margin in both groups and the overall average gap observable between natives and immigrants is actually narrowed by 10% to 12% owing to the higher proportion of men among immigrants. Similarly, the higher proportion of natives in the public sector reduces the observable wage differential. Firm size does not contribute at all to explaining the wage differential within the Oaxaca-Blinder framework, since the two groups are distributed in a very similar way across firm sizes.

4.3. Sample-Selection Correction

Immigrant and native populations do not necessarily have the same (average) employment status. In particular, if the decision to start working is not random between the two analyzed groups, selection in the labor force would yield biased estimates. Factors such as German-language proficiency, education, gender, etc. may differently affect the labor-force

Table 6: Components of the decomposition

	Specification 1		Specification 2		Specification 3	
	Explained (pp)	Explained share of total gap	Explained (pp)	Explained share of total gap	Explained (pp)	Explained share of total gap
Gender	-0.0115	-11.9%	-0.0109	-11.2%	-0.0098	-10.1%
Children	-0.0005	-0.5%	-0.0010	-1.1%	-0.0006	-0.6%
Hours worked	0.0016	1.6%	0.0015	1.6%	0.0020	2.0%
Experience	0.0250	25.7%	0.0268	27.6%	0.0266	27.3%
Education	0.0121	12.4%	0.0086	8.8%	0.0070	7.2%
Occupation	0.0316	32.5%	0.0241	24.8%	0.0078	8.0%
Sector	-0.0077	-7.9%	-0.0074	-7.6%	-0.0069	-7.0%
Firm size	0.0005	0.6%	0.0005	0.5%	0.0005	0.6%
Literacy			0.0424	43.6%	0.0337	34.7%
Writing tasks					0.0083	8.5%
Reading tasks					0.0165	17.0%
Numerical tasks					0.0073	7.5%
Explained gap	0.0512	52.6%	0.0846	87.0%	0.0925	95.1%
Unexplained gap	0.0461	47.4%	0.0127	13.0%	0.0048	4.9%
Total gap	0.0973	100.0%	0.0973	100.0%	0.0973	100.0%

participation of the immigrant population, while at the same time influencing the wages. To avoid this potential problem, we will also compare our previous findings to results corrected for sample selection. In order to identify the selection equation, it is necessary to find instruments which satisfy the exclusion restriction, that is, which affect the decision to work but do not affect the wages.

For a given subsample, the full model should consist of a wage equation and a selection equation, employing a standard latent variables specification:

$$\begin{aligned} y_i^* &= x_i\beta + \varepsilon_i \\ d_i^* &= z_i\gamma + \upsilon_i \\ d_i &= \begin{cases} 1 & \text{if} \quad d_i^* > 0, \\ 0 & \text{otherwise} \end{cases} \\ y_i &= y_i^* \cdot d_i \end{aligned} \tag{4}$$

Household and country-of-origin variables are used to identify the selection equation. Household variables include: gender, age, age squared and the number of children in the household. Variables describing the country of origin are also included: GDP per capita, average labor-force participation and the share of Muslim population. The latter variable is of particular importance for identifying cultural circumstances affecting the labor-force participation of female immigrants.

The probit model for the selection equation is reported in Table 17 in the Appendix. For both immigrants and natives, males are more likely to be working. The effect of age is nonlinear, and follows an inverse U-shaped curve. People with children are less likely to work compared with childless people, and the effect increases with an increasing number of children. Finally, variables describing the country of origin correlate with employment: positively for the case of GDP per capita and negatively for the share of Muslim population.

Table 7 summarizes the OB results corrected for selection bias. We observe an adjusted wage differential of around 11 log points, which, though varying by specification, is higher by one to two log points compared to the raw uncorrected gap.

Table 7: Oaxaca-Blinder decomposition with sample selection correction

	Specification 1		Specification 2		Specification 3		Interaction	
Adj. difference	0.1146	(1.40)	0.1070	(1.29)	0.1140	(1.39)	0.1069	(1.20)
Adj. explained	0.0513**	(2.35)	0.0828 ***	(3.59)	0.0909 ***	(3.88)	0.0891 ***	(3.85)
Adj. unexplained	0.0634	(0.79)	0.0242	(0.30)	0.0231	(0.29)	0.0178	(0.20)

The results for the detailed decomposition of the explained part are summarized in Table 8 (specifications 1-3). The relative contribution of individual variables (as a percentage of the explained part) is shown to be very similar to the previous results presented above, while the explained part itself is lower (in the table, the contribution is represented as a percentage of the total gap) compared to the uncorrected decomposition. Nevertheless, our main findings concerning the importance of literacy skills in explaining the wage gap remain intact: controlling for literacy (specification 2) reduces the unexplained part by almost four log points from the previous 6.3 log points (specification 1).

4.4. Sensitivity Analysis

4.4.1. Restricted Age Group

As already mentioned above, Hofer et al. (2017) find considerably higher raw wage gaps of 15.2 and 18.1 log points for men and women respectively (results for first-generation immigrants), compared to 9.7 log points for the case of our sample. This discrepancy is mainly due

Table 8: Components of the decomposition –corrected for sample selection bias; original sample

	Specification 1		Specification 2		Specification 3	
	Explained (pp)	Explained share of total gap	Explained (pp)	Explained share of total gap	Explained (pp)	Explained share of total gap
Gender	-0.0108	-9.4%	-0.0104	-9.7%	-0.0094	-8.3%
Children	-0.0007	-0.6%	-0.0012	-1.1%	-0.0007	-0.6%
Hours worked	0.0016	1.4%	0.0016	1.5%	0.0020	1.7%
Experience	0.0275	24.0%	0.0287	26.8%	0.0280	24.6%
Education	0.0104	9.0%	0.0074	7.0%	0.0061	5.4%
Occupation	0.0305	26.6%	0.0236	22.1%	0.0078	6.8%
Sector	-0.0077	-6.7%	-0.0074	-6.9%	-0.0069	-6.0%
Firm size	0.0005	0.5%	0.0005	0.4%	0.0005	0.5%
Literacy			0.0401	37.4%	0.0321	28.1%
Writing tasks					0.0077	6.7%
Reading tasks					0.0166	14.5%
Numerical tasks					0.0071	6.3%
Adj. explained	0.0513	44.7%	0.0828	77.4%	0.0909	79.7%
Adj. unexplained	0.0634	55.3%	0.0242	22.6%	0.0231	20.3%
Adj. total	0.1146	100.0%	0.1070	100.0%	0.1140	100.0%

to a difference in observed age groups; while their paper only covers persons aged 20 to 55, our sample extends to persons aged 16 to 65. Although we see no reason to exclude individuals older than 55 from the sample, an examination of the results considering only people aged at least 20 might be worthwhile, since around 50% of both natives and immigrants aged 16 to 19 are still attending school or university and are therefore only occupying minor ”student” jobs, while others are fulfilling their compulsory military or community service duties (in our sample only natives, no first-generation immigrants). Therefore, the wage structure in this age group is fundamentally different from that of people aged at least 20.

Accordingly, Tables 9 and 10 show our findings for the restricted age group (2,337 observations). Most notably, the raw wage gap increases to 13.6 log points, leading to a higher unexplained part of the gap in specifications 1-3 as well as in specification 3 extended by the interaction term literacy×occupation. Note, however, that the Oaxaca-Blinder framework does not allow the effects of discrimination and omitted variables to be distinguished; the remaining unexplained part could therefore be due either to important characteristics not considered in our model, or to discriminatory practices against immigrants. Even though the unexplained gaps in specifications 2 and 3 are no longer statistically insignificant, we can still see a considerable reduction by roughly three log points when adding literacy skills, in line with our previous findings.

Table 9: Oaxaca-Blinder decomposition –ages 20 and above

	Specification 1		Specification 2		Specification 3		Interaction	
Difference	0.1358***	(5.15)	0.1358***	(5.15)	0.1358***	(5.15)	0.1358***	(5.15)
Explained	0.0644***	(3.32)	0.0957***	(4.68)	0.1039***	(4.98)	0.1031***	(4.95)
Unexplained	0.0714***	(3.57)	0.0400**	(2.03)	0.0319*	(1.65)	0.0327*	(1.71)

Table 10: Components of the decomposition –ages 20 and above

	Specification 1		Specification 2		Specification 3	
	Explained (pp)	Explained share of total gap	Explained (pp)	Explained share of total gap	Explained (pp)	Explained share of total gap
Gender	-0.0125	-9.2%	-0.0119	-8.8%	-0.0106	-7.8%
Children	0.0002	0.1%	-0.0003	-0.2%	0.0001	0.1%
Hours worked	0.0029	2.2%	0.0029	2.2%	0.0039	2.8%
Experience	0.0296	21.8%	0.0318	23.4%	0.0310	22.9%
Education	0.0038	2.8%	0.0004	0.3%	-0.0018	-1.3%
Occupation	0.0441	32.5%	0.0365	26.9%	0.0189	13.9%
Sector	-0.0049	-3.6%	-0.0047	-3.5%	-0.0044	-3.2%
Firm size	0.0011	0.8%	0.0010	0.8%	0.0011	0.8%
Literacy			0.0401	29.5%	0.0298	22.0%
Writing tasks					0.0082	6.1%
Reading tasks					0.0207	15.3%
Numerical tasks					0.0069	5.1%
Explained gap	0.0644	47.4%	0.0957	70.5%	0.1039	76.5%
Unexplained gap	0.0714	52.6%	0.0400	29.5%	0.0319	23.5%
Total gap	0.1358	100.0%	0.1358	100.0%	0.1358	100.0%

Finally, we again correct for a possible sample selection bias, as summarized in Table 11. At around 11 log points, the adjusted overall gap is shown to be very close to our previously obtained adjusted gap, while the explained part is higher by roughly one log point. Here, controlling for literacy reduces the unexplained part of the pay differential by roughly three log points. The results for the detailed decomposition for specifications 1-3 are shown in Table 18 in the Appendix.

Table 11: Oaxaca-Blinder decomposition with sample selection correction –ages 20 and above

	Specification 1		Specification 2		Specification 3		Interaction	
Difference	0.1006	(1.28)	0.1028	(1.28)	0.1131	(1.42)	0.1091	(1.30)
Explained	0.0644 ***	(3.32)	0.0964 ***	(4.69)	0.1052 ***	(4.99)	0.1040 ***	(4.96)
Unexplained	0.0363	(0.47)	0.0064	(0.08)	0.0078	(0.10)	0.0052	(0.06)

4.4.2. Immigrants from Non-German-Speaking Countries

So far, we have considered the group of all immigrants together, irrespective of the country of origin. However, one might suspect that immigrants from German-speaking countries face a different situation in the Austrian labor market compared to immigrants from other countries. While the PIAAC data set only includes a selected number of nationalities, we can at least identify foreigners from Germany, who represent by far the most important group of German-speaking immigrants in Austria. In the next step, we will therefore exclude all immigrants from Germany when decomposing the wage gap. The results are shown in Table 12.

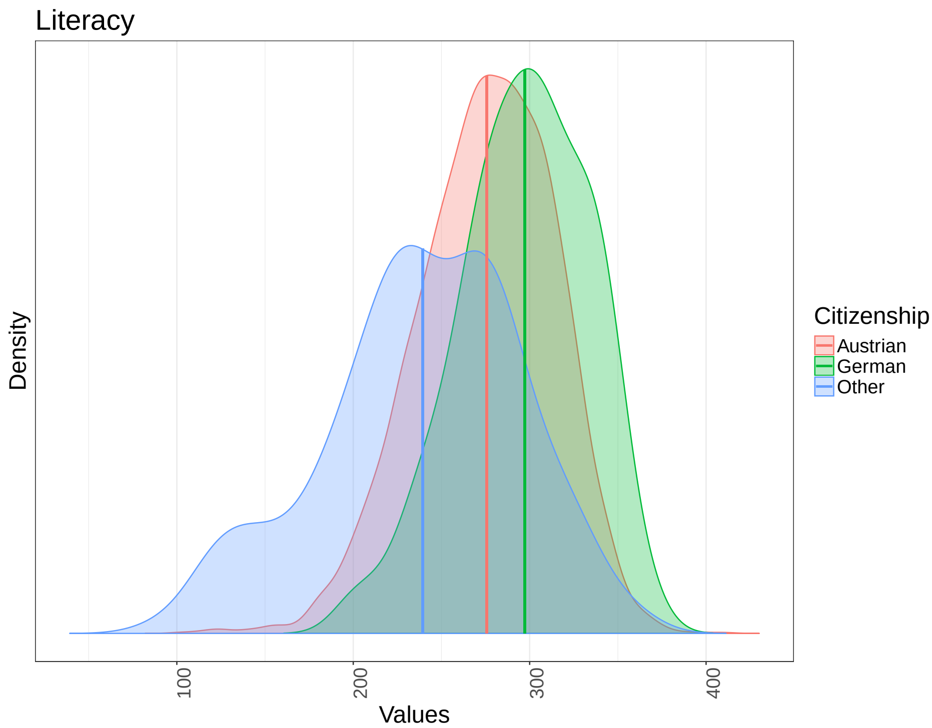
Table 12: Oaxaca-Blinder decompositions without Germans

	Specification 1		Specification 2		Specification 3	
Gender	-0.0112	-7.8%	-0.0106	-7.4%	-0.0095	-6.6%
Children	0.0000	0.0%	-0.0007	-0.5%	-0.0001	0.0%
Hours worked	0.0016	1.1%	0.0016	1.1%	0.0020	1.4%
Experience	0.0312	21.8%	0.0331	23.1%	0.0327	22.8%
Education	0.0261	18.2%	0.0207	14.5%	0.0179	12.5%
Occupation	0.0475	33.2%	0.0374	26.1%	0.0153	10.7%
Sector	-0.0087	-6.0%	-0.0083	-5.8%	-0.0079	-5.5%
Firm size	0.0027	1.9%	0.0024	1.7%	0.0025	1.8%
Literacy			0.0553	38.6%	0.0437	30.5%
Writing tasks					0.0116	8.1%
Reading tasks					0.0228	15.9%
Numerical tasks					0.0099	6.9%
Explained gap	0.0892	62.2%	0.1309	91.3%	0.1411	98.4%
Unexplained gap	0.0542	37.8%	0.0125	8.7%	0.0023	1.6%
Total gap	0.1433	100.0%	0.1433	100.0%	0.1433	100.0%

Compared to the full-sample wage gap of 9.7 log points, the raw gap increases sharply to 14.3 log points when only non-German immigrants are considered. However, the share attributable to differences in characteristics increases for all three specifications. In particular, as one might expect, the part of the wage gap accounted for by differences in literacy skills rises to 4.4 - 5.5 log points (previously 3.4 - 4.2 log points), even though the relative contribution stays around the previous level of roughly one third of the total gap. Indeed, these results suggest that German immigrants earn considerably higher wages than other immigrants, as reflected by the increase in the raw wage gap, and that they also exhibit more favourable characteristics than immigrants from other countries, resulting in a higher share of the raw gap being explainable by differences in characteristic endowments. With regard to literacy proficiency, German immigrants in fact exhibit slightly higher literacy scores than Austrian natives, as shown in Figure 12.

Finally, Table 13 also summarizes the results obtained when accounting for sample selection bias. In this case, the adjusted total gap amounts to about 17.2 log points, with around

Figure 3: Empirical distribution of literacy skills: differences between natives, German-speaking immigrants and other immigrants in Austria



4.2 - 5.2 log points explained by differences in literacy skills. This is very similar to the previous level.

Overall, we can say that excluding German immigrants from the sample results in expected effects, again confirming the significance of literacy proficiency as a source of differences between wages, as has been suggested by our previous findings.

4.4.3. Ñopo Decomposition

The Ñopo (2008) decomposition method typically requires a large sample, as it is based on perfect matching. Given a limited number of observations, we consider the following calculations as a robustness check for our main Oaxaca-Blinder specifications. This additionally allows us to look from a different perspective at the heterogeneity between the groups of natives and immigrants with respect to their endowments of characteristics.

When choosing characteristics to be considered in the Ñopo decomposition, one faces a trade-off. On the one hand, a higher number of variables ensures that matched immigrants

Table 13: Oaxaca-Blinder decompositions without Germans –sample selection correction

	Specification 1		Specification 2		Specification 3	
Gender	-0.0106	-6.2%	-0.0101	-6.3%	-0.0092	-5.3%
Children	-0.0003	-0.2%	-0.0008	-0.5%	-0.0002	-0.1%
Hours worked	0.0016	0.9%	0.0016	1.0%	0.0020	1.2%
Experience	0.0329	19.2%	0.0343	21.3%	0.0337	19.6%
Education	0.0237	13.8%	0.0192	11.9%	0.0168	9.8%
Occupation	0.0461	26.8%	0.0369	22.9%	0.0153	8.9%
Sector	-0.0086	-5.0%	-0.0083	-5.2%	-0.0079	-4.6%
Firm size	0.0027	1.5%	0.0024	1.5%	0.0025	1.5%
Literacy			0.0521	32.3%	0.0415	24.1%
Writing tasks					0.0108	6.3%
Reading tasks					0.0229	13.4%
Numerical tasks					0.0096	5.6%
Adj. explained	0.0875	51.0%	0.1272	78.9%	0.1380	80.3%
Adj. unexplained	0.0842	49.0%	0.0341	21.1%	0.0339	19.7%
Adj. total	0.1717	100.0%	0.1613	100.0%	0.1718	100.0%

and natives are indeed similar with respect to their characteristics, i.e., the comparison of their wages is adequate. On the other hand, however, matching by more variables makes it more difficult to find native matches for each immigrant, which reduces the number of natives and immigrants on the common support that can be compared to each other. We will therefore consider the following three specifications, attempting to balance these two aspects:

- Specification (a) corresponds to our wage regression model employed in the OB decomposition, matching by gender, children, hours worked, experience, education, occupation, firm size and literacy skills.

- Specification (b) includes gender, experience, education, occupation, firm size and literacy skills.

- Specification (c) includes gender, experience, education, occupation and literacy skills.

Furthermore, the literacy skills and experience variables are merged into quintiles, while hours worked per week are reduced to full-time or part-time in order to ensure a sufficient number of matches.

Table 14 summarizes the results for all three specifications. The left two columns correspond to the full sample, while the right two columns only consider individuals aged at least 20. Differences are expressed as a percentage of the average hourly wage of an immigrant[11].

[11]Note that the overall gap Δ varies marginally within columns due to slight changes in the sample size

Specification (a) takes into account the same variables as the OB specifications 1 and 2. Indeed, we can see striking differences between the supports of natives and immigrants, a fact that the standard Oaxaca-Blinder framework fails to reveal. Outside the common support, it could be problematic to relate differences in wages to differences in characteristics.

For OB specification 1, only 54.5% of immigrants can be perfectly matched to 22.6% of natives, making a direct comparison possible only for these subsamples. When including literacy skills also (OB specification 2), only 29.3% of immigrants and 7.1% of natives remain. This further highlights the large differences in literacy skills between the two groups.

Additionally, we can see that the on-support wage gap ($\Delta_0 + \Delta_X$) is substantially smaller than the overall raw gap. The reason is that individuals with extreme values for their characteristics and therefore often also extreme values for their wages, are especially likely to fall out of the common support, with natives presumably tending to exhibit extremes in a favourable direction, and immigrants in an unfavourable direction. Consider the variable capturing education: immigrants with extremely low levels of education (and presumably also low wages) and natives with very high levels of education (and presumably high wages) tend to find a match less easily and therefore fall out of the common support, leading to a lower wage gap in the common support.

Reading Table 14 from specification (c) at the bottom to specification (a) at the top, we can see how increasing the number of included variables leads to a decrease in the common support that concentrates matched natives and immigrants towards the center of their joint characteristics distribution, thereby progressively reducing Δ_0 and Δ_X, the unexplained and the explained parts of the gap on the common support. Accordingly, at the same time, we observe that $\Delta_N + \Delta_M$ increases, indicating that the wage gap is driven by those natives and immigrants for whom no match can be found.

In contrast to the effect of increases in the number of variables within columns, including the literacy variable results in Δ_X (the explained part) being notably higher, in spite of a reduction in the common support. The effect of a lower overall gap on the common support

caused by missing values in some of the variables.

and therefore probably also a lower Δ_X, is more than outweighed by the explanatory value of literacy proficiency for specifications (a) and (b). We see no meaningful difference in the overall gap on the common support for specification (c), but still see a higher value for Δ_X. The Ñopo decomposition therefore confirms the importance of literacy skills for explaining the wage gap, as has already been suggested by the findings from the OB decomposition.

Considering the full sample, controlling for literacy skills can increase the explained part of the gap *on the common support* by roughly between 1.2 and 2.7 percentage points, while we also see a meaningful reduction of Δ_0 (even into the negative regime). When the sample is restricted to individuals aged at least 20, the effect becomes even more pronounced.

Table 14: Ñopo-decomposition

		Full sample		**Age≥20**	
		Without literacy	With literacy	Without literacy	With literacy
Specification (a)	D	10.37%	10.37%	12.67%	12.67%
	D0	3.06%	-6.67%	4.45%	-5.18%
	DN	0.65%	4.22%	0.26%	1.35%
	DM	6.75%	11.71%	8.13%	14.63%
	DX	-0.09%	1.11%	-0.16%	1.87%
	matched nat.	*22.6%*	*7.1%*	*22.6%*	*6.9%*
	matched mig.	*54.5%*	*29.3%*	*53.4%*	*28.0%*
Specification (b)	D	10.48%	10.48%	12.79%	12.79%
	D0	4.14%	-0.26%	5.90%	1.66%
	DN	3.51%	5.70%	3.08%	3.11%
	DM	2.80%	2.33%	3.53%	4.25%
	DX	0.02%	2.71%	0.28%	3.76%
	matched nat.	*42.6%*	*14.9%*	*42.1%*	*14.4%*
	matched mig.	*76.2%*	*47.3%*	*75.0%*	*45.3%*
Specification (c)	D	10.72%	10.72%	13.05%	13.05%
	D0	6.94%	4.79%	9.23%	6.82%
	DN	0.60%	-0.01%	0.22%	-1.18%
	DM	1.92%	2.39%	2.45%	3.51%
	DX	1.27%	3.55%	1.15%	3.90%
	matched nat.	*79.3%*	*38.8%*	*79.0%*	*38.5%*
	matched mig.	*89.6%*	*73.0%*	*88.6%*	*71.1%*

5. Conclusion

While other studies conducted on larger data sets were able to examine the wage differential along the whole wage distribution in Austria (Hofer et al. 2017), the PIAAC data set is limited by the sample size and only permits an analysis of the means. At the same time, however, it allows for the broadening of the current state of research to include a different direction. In contrast to previous studies, we can directly control for literacy skills, an aspect of human capital that plays a major role in an individual's productivity.

Accordingly, we observe that the literacy scores of immigrants are substantially lower than those of natives. Furthermore, we find that literacy skills have a significant impact on wages, but are rewarded differently for natives and immigrants, which, according to our results, seems to be driven by segregation into different occupations. Returns on literacy proficiency for immigrants only become visible when distinguishing between various groups of occupations: higher literacy skills are rewarded in most of the white-collar occupations, but there is no evidence of such effects for other occupations. This indicates that for certain white-collar jobs, improvement of immigrants' literacy skills could meaningfully enhance their wages.

Turning to the decomposition of the wage differential, we can verify that aspects of human capital such as work experience and education previously employed in decompositions by Hofer et al. (2017) indeed play an important role in explaining the immigrant wage gap in Austria. When considering our full sample, taking literacy proficiency into account, we are able to close the unexplained gap almost completely: 87% to 95% (specifications 2 and 3) of the overall gap can be explained. Differences in literacy skills can explain more than a third of the raw wage gap.

In a further step, we also account for a potential sample selection bias, since selection into the labor force might differ between immigrants and natives. After controlling for sample selection, the (adjusted) wage gap rises to around 11 log points. The unexplained part of the wage gap is close to two log points but is not statistically significant and the literacy skills again explain more than three log points of the wage gap.

We also test the sensitivity of our results by changing the age group of our sample, leaving out workers under the age of 20. In the second step, we exclude German immigrants from our sample, since they demonstrate on average better literacy skills than natives. In all of these sensitivity tests, literacy skills consistently reduce the unexplained wage gap by roughly three to four log points. The importance of literacy proficiency in explaining wage differences between immigrants and natives is also confirmed by a matching approach (Ñopo decomposition) conducted as a check of robustness. This approach also reveals considerable differences in the support of the empirical characteristics distributions of natives and immigrants.

To summarize, our findings highlight the importance of literacy skills on wage differences in the Austrian labor market. Literacy skills explain more than three log points of the wage differences between immigrants and natives. Our findings suggest that unexplained pay differences between immigrants and natives are close to zero, but we do not draw any conclusions concerning job discrimination.

Future research related to the immigrant wage gap in Austria could therefore try to simultaneously assess job and pay discrimination within one common framework. Furthermore, examination of the impact of literacy skills on the immigrant wage gap along the whole wage distribution, as well as by gender, would be highly interesting, although to the best of our knowledge, no suitable data set is available at present.

References

Aldashev, A., J. Gernandt, and S. L. Thomsen (2012). The immigrant-native wage gap in germany. *Jahrbücher für Nationalökonomie und Statistik 232*(5), 490–517.

Autor, D. H. and M. J. Handel (2013). Putting tasks to the test: Human capital, job tasks, and wages. *Journal of Labor Economics 31*(S1), S59–S96.

Bartolucci, C. (2014). Understanding the native–immigrant wage gap using matched employer-employee data: evidence from germany. *ILR Review 67*(4), 1166–1202.

Blinder, A. S. (1973). Wage discrimination: reduced form and structural estimates. *Journal of Human resources*, 436–455.

Cai, L. and A. Y. Liu (2015). Wage differentials between immigrants and the native-born in australia. *International Journal of Manpower 36*(3), 374–396.

Chernozhukov, V., I. Fernández-Val, and B. Melly (2013). Inference on counterfactual distributions. *Econometrica 81*(6), 2205–2268.

Chiswick, B. R. (1978). The effect of americanization on the earnings of foreign-born men. *Journal of political Economy 86*(5), 897–921.

Chiswick, B. R. and P. W. Miller (2008). Why is the payoff to schooling smaller for immigrants? *Labour Economics 15*(6), 1317–1340.

Coulombe, S., G. Grenier, and S. Nadeau (2014). Human capital quality and the immigrant wage gap. *IZA Journal of Migration 3*(1), 14.

Dell'Aringa, C., C. Lucifora, and L. Pagani (2015). Earnings differentials between immigrants and natives: the role of occupational attainment. *IZA Journal of Migration 4*(1), 8.

Derwing, T. M. and E. Waugh (2012). Language skills and the social integration of canada's adult immigrants. *IRPP study* (31), 1.

DiNardo, J., N. M. Fortin, and T. Lemieux (1995). Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. Technical report, National bureau of economic research.

Dustmann, C. (1993). Earnings adjustment of temporary migrants. *Journal of Population Economics 6*(2), 153–168.

Fortin, N., T. Lemieux, and S. Firpo (2011). Decomposition methods in economics. *Handbook of labor economics 4*, 1–102.

Garcia, L. M., H. Ñopo, and P. Salardi (2009). Gender and racial wage gaps in brazil 1996-2006: evidence using a matching comparisons approach.

Grandner, T. and D. Gstach (2015). Decomposing wage discrimination in Germany and Austria with counterfactual densities. *Empirica 42*(1), 49–76.

Hanushek, E. A., G. Schwerdt, S. Wiederhold, and L. Woessmann (2015). Returns to skills around the world: Evidence from piaac. *European Economic Review 73*, 103–130.

Heckman, J. J., H. Ichimura, and P. E. Todd (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The review of economic studies 64*(4), 605–654.

Hellerstein, J. K. and D. Neumark (1999). Sex, wages, and productivity: An empirical analysis of israeli firm-level data. *International Economic Review 40*(1), 95–123.

Hofer, H., G. Titelbach, R. Winter-Ebmer, and A. Ahammer (2017). Wage discrimination against immigrants in austria? *Labour 31*(2), 105–126.

Izquierdo, M., A. Lacuesta, and R. Vegas (2009). Assimilation of immigrants in spain: A longitudinal analysis. *Labour Economics 16*(6), 669–678.

Lehmer, F. and J. Ludsteck (2011). The immigrant wage gap in germany: Are east europeans worse off? *International migration review 45*(4), 872–906.

Lehmer, F. and J. Ludsteck (2013). Lohnanpassung von ausländern am deutschen arbeitsmarkt: Das herkunftsland ist von hoher bedeutung. Technical report, IAB-Kurzbericht.

Machado, J. A. and J. Mata (2005). Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of applied Econometrics 20*(4), 445–465.

Nicodemo, C. and R. Ramos (2012). Wage differentials between native and immigrant women in spain: Accounting for differences in support. *International Journal of Manpower 33*(1), 118–136.

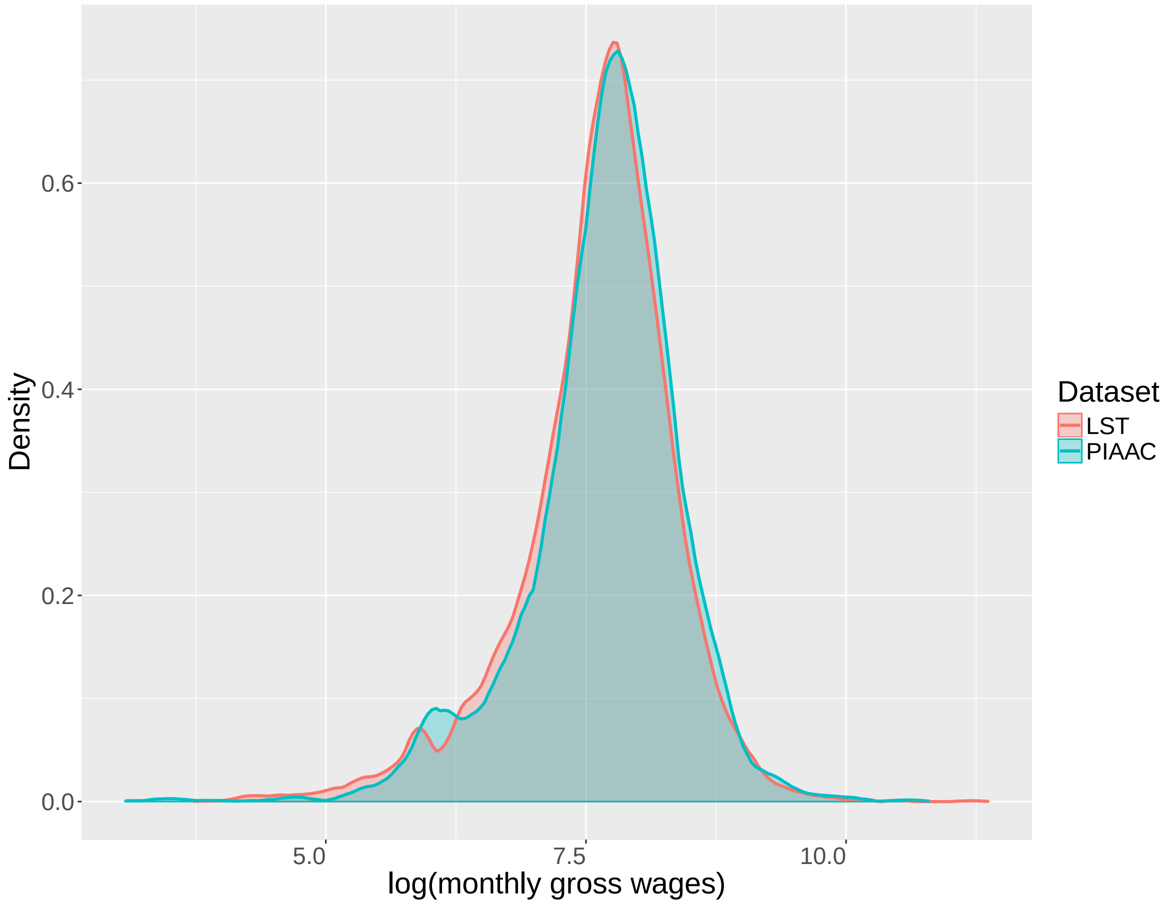
Ñopo, H. (2008). Matching as a tool to decompose wage gaps. *The review of economics and statistics 90*(2), 290–299.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International economic review*, 693–709.

OECD (2016). The survey of adult skills: Readers's companion, second edition.

Appendix A

Figure 4: Distributions of earnings in the wage-tax statistics (LST, red) and the PIAAC (blue) data set



Comment: To make both data sets comparable, we adjust the wage statistics from 2012 for age groups below 65 and include only taxpayers that were employed during the whole year, since PIAAC data are based on hourly wages whereas the tax statistics are on a yearly basis. For the PIAAC data from 2011/2012, we leave out the self-employed, who are not included in the tax statistics.

Table 15: Variables used in the log-linear wage regressions

Variable	**Description**
Male	Dummy variable which takes the value 1 for males
Children	Dummy variable taking the value 1 if the person has at least 1 child
Experience	Years of work experience; quadratic trend also included
Hours worked	Working hours per week
ISCED	Categorical variable for education according to the ISCED 97 classification, taking the levels ISCED1 (primary), ISCED2 (lower secondary), ISCED3 (upper secondary), ISCED4 (post-secondary, non-tertiary), ISCED56 (tertiary, combining ISCED5 and ISCED6)
ISCO1C	Categorical variable for occupation, taking the levels 0 (armed forces), 1 (legislators, senior officials and managers), 2 (professionals), 3 (technicians and associate professionals), 4 (clerks), 5 (service workers and shop and market sales workers), 6 (skilled agricultural and fishery workers), 7 (craft and related trades workers), 8 (plant and machine operators and assemblers) and 9 (elementary occupations)
Sector	Categorical variable taking the values private, public and NGO
Firm size	Categorical variable indicating the number of employees - levels are "up to 10", "11 to 50", "51 to 250", "251 to 1,000" and "more than 1,000"
Literacy	Indicating literacy skills on a scale from 0 (minimum) to 500 (maximum), in the regression scaled by the factor $\frac{1}{100}$
Reading tasks	Indicating the frequency of reading tasks at work; quintiles of an index constructed from the respondents' answers
Writing tasks	Indicating the frequency of writing tasks at work; quintiles of an index constructed from the respondents' answers
Numerical tasks	Indicating the frequency of numerical tasks at work; quintiles of an index constructed from the respondents' answers
Immigrant	Dummy variable taking the value one for individuals belonging to the group of first-generation immigrants, and 0 for natives

Table 16: Components of the decomposition –interaction between literacy skills and occupation

	Interaction	
	Explained (pp)	Explained share of total gap
Gender	-0.0099	-10.1%
Children	-0.0004	-0.4%
Hours worked	0.0020	2.0%
Experience	0.0269	27.6%
Education	0.0078	8.0%
Sector	-0.0064	-6.6%
Firm size	0.0006	0.6%
Occupation $\times$ literacy	0.0383	39.4%
Writing tasks	0.0079	8.1%
Reading tasks	0.0164	16.8%
Numerical tasks	0.0073	7.5%
Explained gap	0.0904	92.9%
Unexplained gap	0.0069	7.1%
Total gap	0.0973	100.0%

Table 17: Selection equation

Age	0.1947***
	(20.19)
Age^2	-0.0026***
	(-22.49)
One child	-0.0513
	(-0.80)
Two children	0.0026
	(0.04)
Three children	-0.1265*
	(-1.69)
Four or more children	-0.3382***
	(-3.32)
GDP per capita	9.13e-06
	(1.61)
LFP in country	-0.0216
	(-1.21)
Share of Muslims	-0.0079**
	(-2.05)
Male	0.0968**
	(2.48)
Constant	-1.5503
	(-1.29)
Observations	4710

Robust z-statistics in parentheses; significance *0.1 **0.05 ***0.01

Table 18: Components of the decomposition corrected for sample selection bias –ages 20 and above

	Specification 1		Specification 2		Specification 3	
	Explained (pp)	Explained share of total gap	Explained (pp)	Explained share of total gap	Explained (pp)	Explained share of total gap
Gender	-0.0125	-12.5%	-0.0121	-11.8%	-0.0109	-9.6%
Children	0.0002	0.2%	-0.0002	-0.2%	0.0003	0.2%
Hours worked	0.0029	2.9%	0.0029	2.8%	0.0038	3.4%
Experience	0.0296	29.4%	0.0311	30.3%	0.0299	26.5%
Education	0.0038	3.8%	0.0005	0.5%	-0.0016	-1.4%
Occupation	0.0442	43.9%	0.0369	35.9%	0.0193	17.1%
Sector	-0.0049	-4.9%	-0.0047	-4.6%	-0.0043	-3.8%
Firm size	0.0011	1.1%	0.0010	1.0%	0.0011	1.0%
Literacy			0.0409	39.8%	0.0311	27.5%
Writing tasks					0.0086	7.6%
Reading tasks					0.0208	18.4%
Numerical tasks					0.0071	6.2%
Adj. explained	0.0644	64.0%	0.0964	93.8%	0.1052	93.1%
Adj. unexplained	0.0363	36.0%	0.0064	6.2%	0.0078	6.9%
Adj. total	0.1006	100.0%	0.1028	100.0%	0.1131	100.0%