Empirica (2019) 46:691–711
https://doi.org/10.1007/s10663-019-09459-0

ORIGINAL PAPER

The wage premium from foreign language skills

Jacek Liwiński[1]

Published online: 21 September 2019

Abstract
The globalisation of labour markets makes language skills one of the key competences required by employers nowadays. The purpose of this paper is to estimate the wage premium from the foreign language skills earned by Poles. Poles seem to be a good case study, because the Polish language is not used for international communication and as many as 58% of Poles command at least one foreign language. I use data from three waves (2012–2014) of the Human Capital Balance survey with a pooled sample of about 35,000 individuals. I estimate a wage equation with the Heckman correction for selection to employment and I check for robustness with propensity score matching. I find that the advanced command of a foreign language yields a wage premium of 11% on average. Interestingly, I find a much higher wage premium from proficiency in Spanish (32%), French (22%) or Italian (15%) than from proficiency in English (11%) or German (12%).

Keywords Foreign language skills · Wage premium · Wage equation

JEL Classification I26 · J24

1 Introduction

Increasing globalisation, growing international trade, foreign direct investment and international labour migration are all causing the demand for language competences to grow, and according to forecasts, this trend will continue in the future (Antonietti and Loi 2014; Isphording 2015). In Central and Eastern European countries, including Poland, these changes intensified after accession to the European Union (EU). Additionally, a significant increase in the inflow of foreign direct investment

This paper has been presented at the MIRDEC 5th Conference on Social Science (Vienna, 12–14.09.2017).

✉ Jacek Liwiński
jliwinski@wne.uw.edu.pl

[1] Faculty of Economic Sciences, University of Warsaw, ul. Dluga 44/50, 00-241 Warsaw, Poland

to Poland has been observed since 2004, and approximately 900,000 Poles have migrated to the United Kingdom and Ireland with the opening of EU labour markets in 2006. Under such circumstances, the command of foreign languages is becoming one of the key competences on the global labour market. Hence, it is not surprising that substantial public and private funds are allocated to language education. Learning two foreign languages is obligatory in schools in Poland, which is similar to most European countries. Furthermore, private language schools are thriving in Poland. In the European context, language skills are promoted by initiatives such as the Erasmus programme, which is financed from public funds and individual financial contributions. As language education can be expensive, it seems interesting from an economic point of view whether this type of education brings any benefits—and in particular, wage benefits—to individuals in the labour market.

Our purpose is to estimate the wage premium from foreign language skills (FLS) that is obtained by Poles. Poles seem to be a good case for this type of analysis because the Polish language is not used for international communication and as many as 58% of Poles declare the command of at least one foreign language.[1] I estimated an extended Mincer wage equation using OLS. The analysis was based on data from three waves (2012–2014) of the Human Capital Balance survey with a pooled sample of about 35,000 individuals.

This study contributes to the literature on economic returns to FLS in the following ways. First, this is one of the few studies that identifies the wage premium from proficiency in many different languages—not only English (Di Paolo and Tansel 2015; Williams 2011). Second, my study is one of only a few examining Central and Eastern European countries (Toomet 2011; Fabo et al. 2017) and is the first for Poland.

This paper is organised into seven sections. In the first section, I present an overview of the literature on the impact of language competences on wages. The second and the third sections present the data and method of analysis. The fourth section contains the descriptive statistics of the sample and the fifth discusses the results of the empirical analysis. The robustness analysis is presented in the sixth section, and at the end of the paper, the most important findings are summarised.

2 Review of literature

The relevant literature discusses four reasons why proficiency in foreign languages should influence earnings. First, language skills can have a positive impact on productivity because they enhance the effectiveness of communication at work. This includes both internal communication—with peers or managers—and external communication—with clients or suppliers (Ku and Zussmann 2010). Second, learning a foreign language, and then using it, can translate into improved cognitive skills and, consequently, greater productivity at work. Adesope et al. (2010) show that bilinguals have a cognitive advantage over monolinguals in executive functions involving

[1] Own analysis based on Human Capital Balance survey data (2012–2014).

mental flexibility, inhibitory control, attention control and task switching as well as creativity, flexibility and originality in problem solving (Di Paolo and Tansel 2015). Hence, the above two channels of the impact of language skills on wages correspond to the human capital theory, as they perceive the increments of skills to be the source of higher earnings. The third explanation refers to the signalling theory (Spence 1973). Based on the signalling theory, one can argue that the command of a foreign language may be a signal of greater abilities and cognitive skills, thereby suggesting higher potential productivity. And fourth, the command of a foreign language may open doors to more prestigious professions that are also likely to be better remunerated (Chiswick and Miller 2009). This is confirmed by Chiswick and Miller (2010) who find that workers obtain a premium if their language skills match linguistic requirements in the workplace. Moreover, Aldashev et al. (2009) suggest that the positive effect of language proficiency among immigrants in Germany is entirely driven by occupational selection, as the wage premium disappears once controlling for the endogenous decision to enter a particular economic sector and occupation.

There are three strands of research on the wage premium from language skills. The first and most popular examines the wage premium earned by immigrants in the host country [see Dustmann (1994), Dustmann and van Soest (2001, 2002), Stöhr (2015) for Germany; Berman et al. (2003), Lang and Siniver (2009) for Israel; Isphording (2015) for Spain; Leslie and Lindley (2001), Shields and Price (2002), Dustmann and Fabbri (2003) for the United Kingdom; and Bleakley and Chin (2010) for the United States]. It is well established in this literature that immigrants with host country language skills obtain a positive wage premium.

The second strand refers to the wage premium from bilingualism obtained by natives in multilingual labour markets [see Shapiro and Stelcner (1997), Albouy (2008) for Canada; Henley and Jones (2005) for Wales; Grin and Sfreddo (1998), Cattaneo and Winkelmann (2005) for Switzerland; and Rendon (2007), Di Paolo (2011), Di Paolo and Raymond (2012) for Catalonia in Spain]. The results obtained usually show that proficiency in a second language is remunerated in the labour market.

The third strand, to which my paper is most closely related, focuses on the wage premium from the command of foreign languages earned by natives. Most of these papers focus on developed countries [see Grin (2001) for Switzerland; Fry and Lowell (2003), Saiz and Zoido (2005) for the United States; Lang and Siniver (2009) for Israel; Ginsburgh and Prieto-Rodriguez (2011, 2013), Williams (2011) for a comparative analysis of several Western European countries; and Stöhr (2015) for Germany]. Other studies examine developing countries [see Levinsohn (2007) and Casale and Posel (2011) for South Africa; Azam et al. (2013) for India; Di Paolo and Tansel (2015) for Turkey; and Guo and Sun (2014), Wang et al. (2017) for China]. Thus far, not much research in this strand has been conducted for Central and Eastern European countries [see Toomet (2011) for Latvia and Estonia and Fabo et al. (2017) for Czechia, Slovakia and Hungary]. Furthermore, most of these studies concentrate on the command of English, with only a few analysing premiums from other languages [see Ginsburgh and Prieto-Rodriguez (2011), Williams (2011), Di Paolo and Tansel (2015)]. The literature in this strand finds a positive wage premium

from the command of English, while in some countries—from other foreign languages as well.

According to my best knowledge, there are no studies for Central and Eastern European countries on the wage premium from proficiency in any other foreign languages than English. My research is intended to fill this gap in the literature.

3 Data

The analysis is based on the Human Capital Balance survey for the period of 2012–2014.[2] The survey is a unique source of unit data about the human capital resources of Poles. It also provides information about the situation of individuals in the labour market. The sample is representative for the working age population, i.e. women aged 18–59 and men aged 18–64, residing in Poland. The sample consists of approximately 17,600 individuals in each yearly wave of the survey.

For this study, it is important that the survey included questions about FLS. Respondents were asked to list all languages they knew and to assess their level of command for the three languages they knew best (using a six-grade scale) in four linguistic competences: speaking, writing, reading and listening comprehension. Although the measure was subjective, it seems that owing to the relatively broad rating scale, the variation of language skills between individuals can be properly identified.

The database also contains information on 12 additional skills, including:

- Searching and analysing information and drawing conclusions;
- Operating, installing, and repairing technical equipment;
- Performing calculations;
- Operating a computer and using the Internet;
- Artistic and creative abilities;
- Physical fitness;
- Self-organisation of work and showing initiative;
- Networking;
- Arranging and performing office tasks;
- Managerial skills and organising other people's work;
- Availability;
- Fluency in spoken and written Polish.

The respondents were asked to assess their skills in the above-listed areas, using a five-grade scale. Hence, I have a map of respondents' skills at my disposal. Although these skills are likely to be affected by the abilities and motivation of the respondents, and may suffer from misreporting as the assessments are subjective, they may

[2] The Human Capital Balance (BKL) survey was conducted in the years 2010–2014 by the Polish Agency for Enterprise Development (PARP) in co-operation with Jagiellonian University.

affect FLS and thus it is worth controlling for them in the wage equation to decrease any possible endogeneity bias.

Importantly, I wanted to exclude all migrants from my sample; however, I had no information on the nationality of the respondents. Consequently, I dropped all individuals who reported Polish as a foreign language. I believe that in this way I removed all individuals for whom Polish is not their native language, because foreigners who do not speak Polish at all (and would not report Polish as a foreign language they know) would not be able to participate in the survey, as the questionnaire is in Polish only.

Additionally, the sample is limited to individuals who were self-employed or employed on a contract within 12 months before the survey and who reported their average earnings from that period. Furthermore, in order to use the Heckman correction, respondents who were unemployed or economically inactive were included in the sample. Eventually, the sample consisted of 35,579 respondents, of which 14,145 were employed.

4 Method

There are at least three possible sources of bias associated with the OLS estimates of the returns to FLS. First, there may be unobserved heterogeneity affecting both FLS and earnings (Chiswick and Miller 1995). Second, there may be reverse causality as individuals earning more can invest more in improving their FLS (Chen et al. 2014). Third, self-reported FLS may suffer from misclassification errors (Dustmann and Van Soest 2001, 2002). Although it seems difficult to eliminate all possible sources of endogeneity, I adopted several approaches to address potential bias.

The first and most important problem is the potential bias resulting from the endogeneity of FLS. If individuals studying foreign languages are on average more able, skilled or motivated, we may expect that the OLS estimate of the wage premium will be overestimated as I do not control for unobservable abilities and skills. The wage premium estimated in this way will reflect not only returns to FLS, but also returns to other skills and abilities that were useful for studying foreign languages and are useful at work. Thus, the standard approach to address unobserved heterogeneity is to use OLS and include control variables that proxy for cognitive skills (Toomet 2011; Fabo et al. 2017; Grin 2001; Fry and Lowell 2003; Ginsburgh and Prieto-Rodriguez 2011; Di Paolo and Tansel 2015). However, the results are rather descriptive in this case and cannot be causally interpreted. Other methods used include: propensity score matching (Saiz and Zoido 2005), panel data methods (Saiz and Zoido 2005; Lang and Siniver 2009) and instrumental variables (Wang et al. 2017). Typically, the wage returns to FLS are identified by using foreign language proficiency as the key independent variable.

As I did not have any variable in my database that could serve as an instrument, I used OLS estimation and included a full set of control variables in the specification to reduce concern that selection may bias the estimates. For this purpose, I used a unique property of the Human Capital Balance survey database—the information on the respondents' 12 additional skills. Considering that the development

of skills—both linguistic and non-linguistic—is determined by abilities and motivation, the inclusion of a wide set of respondent skills to the wage equation may, to a certain degree, reduce the endogeneity bias. However, I treat this approach rather as an exercise and I do not claim that the endogeneity bias will unquestionably be reduced in this way.

The second problem is a potential reverse causality because individuals with higher wages are more likely to afford the payments for private language courses. The share of adults (aged 18–64) in my sample who learned foreign languages within 12 months before the survey is very small (0.4%), while all young people in Poland have to learn two foreign languages at school up to the age of 18. As the share of adults learning foreign languages is marginal, I argue that the estimates should not be substantially biased as a result of reverse causality. Nevertheless, I address this issue by running a robustness check in Sect. 6.

The third problem is the misreporting and rounding of FLS, as this skill is self-assessed by respondents. In order to minimize the bias that may be caused by misreporting, I based my measure of FLS on the key language production skill (i.e. speaking skill) only. Though I had information on reading and listening comprehension, I decided not to use these measures to eliminate possible bias resulting from the overestimation of these skills.[3] Regarding rounding error, I believe that this error is not large because respondents in my sample were asked to assess their FLS on a relatively wide scale that has six levels. Obviously, I am aware that despite my efforts to minimize misreporting and rounding errors, they are present to some extent in my sample and they may bias my estimates.

My identification strategy is to use OLS to estimate the following wage equation:

$$\ln(w_i) = \beta_0 + FLS_i\beta_1 + S_i\beta_2 + X_i\beta_3 + \varepsilon_i \quad (1)$$

where: the dependent variable $\ln(\boldsymbol{w}_i)$ is the natural logarithm of hourly net earnings,[4] $\boldsymbol{FLS}_i$ is a vector of variables representing the level of foreign language skills, $\boldsymbol{S}_i$ covers the map of respondents' skills and vector $\boldsymbol{X}_i$—other factors that may have an impact on earnings.

Vector $\boldsymbol{FLS}_i$ consists of three binary variables representing the following levels of language skills: elementary, intermediate and advanced. I created these variables based on the assessment of the key language production skill—speaking. As the level of this skill was rated by respondents from 1 to 6, I defined FLS as:

- Elementary, if the reported level was 1 or 2;
- Intermediate, if the reported level was 3 or 4;
- Advanced, if the reported level was 5 or 6.

The map of skills ($\boldsymbol{S}_i$) covers the 12 skills listed in the previous section. The model includes a separate variable for each skill, taking values from 1 to 5. Furthermore, the model also includes other control variables ($\boldsymbol{X}_i$) that represent

[3] I gratefully acknowledge that this approach was suggested by an anonymous referee.

[4] The earnings are expressed in 2014 prices.

respondents' features (gender, age, education level), as well as the characteristics of the local labour market (place of residence, region, year of survey). All independent variables are listed in Table 8 in the Appendix.

In order to eliminate outliers, I deleted 0.2% of the upper and lower extreme values from the distribution of the hourly net earnings. The linear regression model was estimated using OLS with heteroscedasticity-robust standard errors.

5 Descriptive statistics

The descriptive statistics of the sample presented in Table 1 show that as many as 57% respondents have at least elementary knowledge of one foreign language. Respondents most often declare the command of English (37%), Russian (20%), and German (19%). The command of Spanish, Italian, or French is rare—as little as 0.7%, 0.9%, and 1.9%, respectively, of the individuals in the sample know these languages.

Table 2 shows that 25% of respondents who declared command of a foreign language rated their language skills as advanced, 60%—as intermediate and 15%—as elementary. An advanced skill level is declared by a relatively high proportion of respondents with a command of Italian (31%) and English (26%) and a relatively low proportion of those with a command of French (13%), Russian (12%), and German (16%).

Table 1 Descriptive statistics of the sample by basic variables. *Source* Author's own analyses based on Human Capital Balance survey data

Variable	*N*	Mean	SD	Min	Max
Hourly net earnings (PLN)	14,118	13.96	10.21	0.89	153.67
Woman	35,579	0.502	0.500	0	1
Age	35,579	38.9	14.0	18	64
Education: lower secondary or below	35,579	0.215	0.411	0	1
Education: basic vocational	35,579	0.296	0.456	0	1
Education: secondary	35,579	0.355	0.479	0	1
Education: tertiary	35,579	0.134	0.341	0	1
Foreign language: any	35,579	0.571	0.495	0	1
Foreign language: English	35,579	0.367	0.482	0	1
Foreign language: French	35,579	0.019	0.138	0	1
Foreign language: German	35,579	0.188	0.391	0	1
Foreign language: Russian	35,579	0.203	0.402	0	1
Foreign language: Italian	35,579	0.009	0.096	0	1
Foreign language: Spanish	35,579	0.007	0.084	0	1

Table 2 Sample distribution by level of foreign language skills. *Source* Author's own analyses based on Human Capital Balance survey data

Foreign language	*N*	Elementary (%)	Intermediate (%)	Advanced (%)
Any	20,292	15.4	59.5	25.1
English	13,001	14.7	59.1	26.2
French	681	32.7	54.5	12.8
German	6643	25.5	58.8	15.7
Russian	7191	26.6	61.2	12.2
Italian	327	15.6	53.2	31.2
Spanish	250	24.0	55.6	20.4

6 Results

The estimations of the wage equation are presented in Table 3. Estimation (1) is unconditional in the sense that hourly earnings are regressed on FLS only. Both advanced and intermediate command of a foreign language are positively correlated with earnings. Yet, these correlations are getting lower as successive explanatory variables are included in the model (specifications 2 and 3). In particular, after the other respondents' skills ($\boldsymbol{S}_i$) are included in the model, the wage premium from advanced FLS falls from 15 to 11% and the premium from intermediate FLS—from 7 to 5%. This decrease may result from a reduction of the endogeneity bias owing to the inclusion of the set of other skills. The reason for this is that respondents' abilities and motivation likely have a positive impact on both linguistic and non-linguistic skills; hence, we can expect a decrease in the estimate of the wage premium to FLS after the set of non-linguistic skills is included in the model.

Knowing that language skills are positively correlated with wages, it is worth finding out which languages yield the highest wage premium. Accordingly, I estimated a model with dummies for three levels of FLS for all of the foreign languages under analysis (English, French, German, Russian, Italian, and Spanish) and the full set of controls (as in specification 3 in Table 3). The results are shown in Table 4. We can see that respondents obtain a wage premium from the advanced command of almost any language covered by the analysis. The premium is particularly high from

Table 3 OLS estimates of the wage equation. *Source* Author's own analyses based on Human Capital Balance survey data

Model specification	(1)	(2)	(3)
Elementary FLS	0.001	−0.011	−0.015
Intermediate FLS	0.144***	0.073***	0.051***
Advanced FLS	0.292***	0.156***	0.115***
Other skills (S_i)	No	No	Yes
Other control variables ($\boldsymbol{X}_i$)	No	Yes	Yes
Number of observations	14,118	14,118	14,118
R2	0.042	0.191	0.212

***/**/* stand for 1%, 5%, and 10% significance, respectively

Table 4 OLS estimates of the wage equation by languages. *Source* Author's own analyses based on Human Capital Balance survey data

Model specification	English	French	German	Russian	Italian	Spanish
Elementary FLS	0.001	−0.005	−0.008	−0.024	−0.221	−0.125
Intermediate FLS	0.055***	0.128***	0.059***	−0.004	−0.071	0.049
Advanced FLS	0.115***	0.223**	0.125***	−0.009	0.154*	0.324***
Other skills ($\boldsymbol{S}_i$)	Yes					
Other control variables ($\boldsymbol{X}_i$)	Yes					
Number of observations	14,118					
R2	0.215					

All estimates come from one specification of the model that includes dummies for three levels of FLS for all the foreign languages shown in the table (English, French, German, Russian, Italian, and Spanish); ***/**/* stand for 1%, 5%, and 10% significance, respectively

the advanced knowledge of Spanish (32%), French (22%), and Italian (15%). The premium from the advanced command of English and German is much lower (11% and 12%, respectively).[5] Such a substantial difference may be due to the fact that the command of the first three languages referred to above is very rare among Poles (these languages are declared by 0.7%, 1.9%, and 0.9% of the sample, respectively), while the last two are much more common (37% and 19%, respectively). Interestingly, the command of Russian does not yield any wage premium at all. It seems that the large investments in teaching this language at schools during the pre-transition period in Poland (learning Russian was mandatory in Polish schools and universities before 1990) do not bring any return because the demand for proficiency in Russian is small in the Polish labour market in relation to the supply.

7 Robustness checks

I have tested the robustness of my results in three ways. First, I estimated the baseline model after dropping the observations for adults who attend language courses. Second, I applied the Heckman correction for selection to employment. Finally, I used propensity score matching (PSM) as an alternative method of estimation.

My first robustness check was to re-estimate the wage equation after dropping the observations for adults who attend language courses because I wanted to find out whether potential reverse causality has any impact on my estimates. Table 5 shows that when I dropped the observations for adults attending language courses the results hardly changed. Thus, it is likely that reverse causality related to participation

[5] As the numbers of individuals speaking French, Italian or Spanish are relatively small in my sample, which may affect the results, I pooled the three Romance languages and estimated the joint wage premium to proficiency in these languages. I found that the wage premium from intermediate FLS is 5.7%, which is basically the same as for English and German, but the wage premium from advanced FLS is as high as 20% and highly statistically significant.

Table 5 Robustness check: reverse causality. *Source* Author's own analyses based on Human Capital Balance survey data

Model specification	(1) Baseline model	(2) No language courses
Elementary FLS	−0.015	−0.015
Intermediate FLS	0.051***	0.050***
Advanced FLS	0.115***	0.115***
Other skills (S_i)	Yes	Yes
Other control variables ($\boldsymbol{X}_i$)	Yes	Yes
Number of observations	14,118	14,056
R2	0.212	0.212

***/**/* stand for 1%, 5%, and 10% significance, respectively

in language courses does not substantially affect my results. However, we cannot completely rule out the reverse causality bias, as it is still possible that individuals earning higher wages have more opportunities to learn foreign languages without attending courses. For example, they may be working in multinational companies and thus exposed to an international environment. Similarly, they may be more likely to travel abroad for leisure, which also provides an opportunity to improve FLS.

Second, as my baseline OLS estimates may be biased due to selection to employment, I applied the correction procedure proposed by Heckman (1979). I used two variables as exclusion restrictions: "having a child aged 0–4 years", "woman having a child aged 0–4 years". I used women having a child as an instrument because in Poland public nurseries for children aged 0–2 essentially do not exist and the availability of kindergartens for children aged 3–4 is limited: in 2013, only 6% of children aged 2 attended nurseries and 59% of children aged 3–4 attended kindergartens (CSO 2013). The availability of kindergartens for children aged 5 is better, because they can attend the pre-primary sections of primary schools; consequently, as many as 92% of children aged 5 attended kindergartens in 2013 (CSO 2013). At the age of 6, children start compulsory pre-school education in Poland. This is why mothers usually do not work and take care of small children up to the age of 2, and some of them continue doing this until the child is 4–5 years old.

The variables to be used as exclusion restrictions should be correlated with the probability of being employed, but not with the outcome variable. I find that both instruments are, as expected, negatively correlated with the probability of being employed and they are not significant when included in the wage regression. Furthermore, the instruments may be regarded as strong since the value of the chi^2 in the Wald test of independence of equations is above 10, and the main and selection equations are independent at 0.001%. Table 6 shows that after applying the Heckman correction, the results are essentially the same as the baseline results.

Third, I used the PSM as an alternative method of estimation. In my study, the basic idea of the PSM method is to match the individuals with FLS (treatment group) with those who have a similar propensity to command a foreign language but do not actually command any (control group). I estimated this propensity based

Table 6 Heckit estimates of the wage equation. *Source* Author's own analyses based on Human Capital Survey data

Model specification	(1)	(2)	(3)
Elementary FLS	0.011	−0.014	−0.014
Intermediate FLS	0.146***	0.076***	0.057***
Advanced FLS	0.283***	0.141***	0.107***
Other skills (S_i)	No	No	Yes
Other control variables ($\boldsymbol{X}_i$)	No	Yes	Yes
Number of observations: selection equation	35,563	35,563	35,563
Number of observations: wage equation	14,110	14,110	14,110
chi2	99.84	11.49	14.07
Prob > chi^2	0	0.0007	0.0002

The selection equation includes two variables: having a child aged 0–4 years, woman having a child aged 0–4 years; ***/**/* stand for 1%, 5%, and 10% significance, respectively

on the observable characteristics of individuals which were included in the wage regression. The wage premium from language skills was estimated as the difference between the mean wage levels in the treatment and control groups.

To identify the propensity to command a foreign language for each individual in the sample, I estimated the following probit model:

$$FL_{ki} = Z_{ki}\gamma_k + \nu_{ki} \tag{2}$$

where the dependent variable ($\boldsymbol{FL}_{ki}$) is equal to 1 if the respondent i commands any foreign language at the level k, and 0 if he or she commands no foreign language at all. The model was estimated for three levels of FLS separately, that is, for $k = \{\text{elementary; intermediate; advanced}\}$. $\boldsymbol{Z}_{ki}$ includes the individual characteristics of respondents, while $\boldsymbol{\nu}_{ki}$ is the random error. As a starting point, I used the variables listed in Table 8 in the Appendix (except for FLS) as explanatory variables, and then I reduced the model by removing, step-by-step, the variables insignificant at 10%.

In order to achieve the best quality of matching as possible, I applied the following matching procedure for each level of FLS. First, I tried the nearest neighbour (NN) matching with 1 to 6 neighbours and looked for the number of neighbours that minimise the Rubin's B statistic, provided that Rubin's R is between 0.5 and 2.[6] In the second step, I tried adding a caliper to NN matching. I tried four calipers: 0.001, 0.01, 0.05, and 0.1. As a result of this step, I chose the method providing the best quality of matching as selected by a low Rubin's B statistic. The last step in the matching procedure was to apply the exact (1:1) matching based on the survey year and on the region (voivodeship). After matching, the average treatment effect on the

[6] Rubin's B is the absolute standardised difference of the means of the linear index of the propensity score in the treated and (matched) non-treated group. Rubin's R is the ratio of the treated to (matched) non-treated variances of the propensity score index. Rubin (2001) recommends that B be less than 25 and that R be between 0.5 and 2 for the samples to be considered sufficiently balanced.

Table 7 PSM estimates of the wage premium from language skills. *Source* Author's own analyses based on Human Capital Balance survey data

	Foreign language skills (baseline category: no foreign language skills)		
	Elementary	Intermediate	Advanced
ATT	−0.017	0.046***	0.096***
Standard error	0.020	0.014	0.025
Neighbours	2	3	3
Caliper	–	0.05	0.01
1:1 matching variables	Survey year, region	Survey year, region	Survey year, region
B statistics (matched)	10.5	13.0	13.8
R statistics (matched)	1.00	1.17	0.96
No. of obs.: untreated	5844	5844	5844
No. of obs.: treated on support	1337	4930	1011
No. of obs.: treated off support	0	111	908

***/**/* stand for 1%, 5%, and 10% significance, respectively

treated (ATT) was estimated as the difference between the mean log of the hourly wage rate in the treatment and control groups. Standard errors were bootstrapped by performing 500 replications.

The PSM estimates presented in Table 7 are similar in qualitative and quantitative terms to the OLS estimates (in Table 4). Individuals enjoy a positive wage premium from both the intermediate (5%) and advanced (10%) command of a foreign language.

8 Conclusions

The analysis presented above shows that, in general, FLS yield a positive wage premium to Poles. Yet, the size of this premium is strongly differentiated by the language. In Poland, a high wage premium can be obtained from the advanced command of Spanish, French, and Italian (32%, 22%, and 15%, respectively), which is most likely due to the fact that the supply of individuals fluent in these languages is low when compared to demand. There are many individuals fluent in English or German in Poland and, presumably, this is why the wage premium is relatively modest with respect to these two languages (11% and 12%, respectively). Fluency in Russian, on the other hand, is not rewarded at all on the Polish labour market. This is likely due to the small demand in relation to the supply.

From a theoretical point of view, these findings may be interpreted using two alternative approaches. According to the human capital theory, the positive

correlation between FLS and earnings may result from the higher productivity of individuals proficient in foreign languages, which is rewarded by the employers demanding these skills. The alternative explanation is based on the signalling hypothesis: proficiency in foreign languages may be a signal of higher abilities and, thus, employers may pay higher wages to individuals with high FLS skills in order to employ those most able, even if FLS are not necessary at work. As my methodology does not allow us to identify the casual effect of learning foreign languages, I cannot claim that FLS have an impact on productivity. Nevertheless, individuals who invest in their language skills can expect to earn higher wages in the future, either due to higher productivity or due to signalling higher abilities. However, in order to assess whether it worth investing in FLS, one would need to compare the total wage gains to the total cost of this investment.

Obviously, my results are subject to a number of limitations. First, both the level of FLS and the amount of earnings are declared by respondents; consequently, they may be subject to various sources of measurement error, such as misreporting or misclassification. Second, I do not claim to identify a causal effect of learning foreign languages by Poles. My results are rather descriptive in the sense that they show a positive correlation between FLS and earnings. Most likely, the results are biased due to unobservable heterogeneity, as neither OLS nor PSM estimation allows for the elimination of this kind of bias completely.

Acknowledgements We thank participants of the MIRDEC 5th Conference on Social Science (Vienna, 12–14.09.2017) as well as two anonymous referees for their useful comments. Nonetheless, the usual disclaimer applies.

Appendix

See Tables 8, 9, and 10.

Table 8 Independent variables in the wage equation. *Source* Author's own elaboration

Independent variable	Value classes
Foreign language skills (FLS_i)	1—elementary; 2—intermediate; 3—advanced; 0*—none
Other skills (S_i)	
Searching and analysing information and drawing conclusions	1—low; 2—basic; 3—medium; 4—high; 5—very high
Operating, installing, and repairing technical equipment	1—low; 2—basic; 3—medium; 4—high; 5—very high
Performing calculations	1—low; 2—basic; 3—medium; 4—high; 5—very high
Operating a computer and using the Internet	1—low; 2—basic; 3—medium; 4—high; 5—very high
Artistic and creative abilities	1—low; 2—basic; 3—medium; 4—high; 5—very high
Physical fitness	1—low; 2—basic; 3—medium; 4—high; 5—very high
Self-organisation of work and showing initiative	1—low; 2—basic; 3—medium; 4—high; 5—very high
Networking	1—low; 2—basic; 3—medium; 4—high; 5—very high
Arranging and performing office tasks	1—low; 2—basic; 3—medium; 4—high; 5—very high
Managerial skills and organising other people's work	1—low; 2—basic; 3—medium; 4—high; 5—very high
Availability	1—low; 2—basic; 3—medium; 4—high; 5—very high
Fluency in spoken and written Polish	1—low; 2—basic; 3—medium; 4—high; 5—very high
Other control variables (Xi)	
Woman	1—woman; 0*—man
Age	Continuous variable
Age squared	Continuous variable
Education	1*—lower secondary or below; 2—basic vocational; 3—secondary; 4—tertiary
Place of residence	1*—rural; 2—town < 10,000 inhabitants; 3—town 10,000-19,999 inhabitants; 4—town 20,000-49,999 inhabitants; 5—town 50,000-99,999 inhabitants; 6—town 100,000-199,999 inhabitants; 7—town 200,000-499,999 inhabitants; 8—town ≥ 500,000 inhabitants; 9—Warsaw
Region (voivodeship)	1*—Dolnośląskie; 2—Kujawsko-pomorskie; 3—Lubelskie; 4—Lubuskie; 5—Łódzkie; 6—Małopolskie; 7—Mazowieckie; 8—Opolskie; 9—Podkarpackie; 10—Podlaskie; 11—Pomorskie; 12—Śląskie; 13—Świętokrzyskie; 14—Warmińsko-mazurskie; 15—Wielkopolskie; 16—Zachodniopomorskie
Survey year	Dummy variables (2012–2014)

Asterisks indicate the base category

Table 9 OLS estimates of the wage equation: full set of estimates. *Source* Author's own analyses based on Human Capital Balance survey data

Model specification	(1)	(2)	(3)
Elementary FLS	0.001	−0.011	−0.015
Intermediate FLS	0.144***	0.073***	0.051***
Advanced FLS	0.292***	0.156***	0.115***
Other skills (S_i)	No	No	Yes
Searching and analysing information and drawing conclusions			0.034***
Operating, installing, and repairing technical equipment			0.005
Performing calculations			0.002
Operating a computer and using the Internet			0.043***
Artistic and creative abilities			−0.001
Physical fitness			0.001
Self-organisation of work and showing initiative			0.004
Networking			−0.011**
Arranging and performing office tasks			−0.008*
Managerial skills and organising other people's work			0.039***
Availability			−0.000
Fluency in spoken and written Polish			−0.011**
Other control variables ($\boldsymbol{X}_i$)	No	Yes	Yes
Woman		−0.198***	−0.182***
Age		0.040***	0.039***
Age squared		−0.000***	−0.000***
Education: lower secondary or below		Ref.	Ref.
Education: basic vocational		0.064***	0.043**
Education: secondary		0.187***	0.115***
Education: tertiary		0.469***	0.352***
Place of residence: rural		Ref.	Ref.
Place of residence: town < 10,000 inhabitants		0.015	0.005
Place of residence: town 10,000–19,999 inhabitants		0.043***	0.038***
Place of residence: town 20,000–49,999 inhabitants		0.080***	0.063***
Place of residence: town 50,000–99,999 inhabitants		0.097***	0.083***
Place of residence: town 100,000–199,999 inhabitants		0.118***	0.093***
Place of residence: town 200,000–499,999 inhabitants		0.111***	0.097***
Place of residence: town ≥ 500,000 inhabitants		0.113***	0.099***
Place of residence: Warsaw		0.217***	0.201***
Region: Dolnośląskie		Ref.	Ref.
Region: Kujawsko-pomorskie		−0.064***	−0.057***
Region: Lubelskie		−0.092***	−0.078***
Region: Lubuskie		0.025	0.041**
Region: Łódzkie		−0.081***	−0.069***
Region: Małopolskie		0.014	0.021
Region: Mazowieckie		0.049**	0.040*
Region: Opolskie		0.045**	0.063***

Table 9 (continued)

Model specification	(1)	(2)	(3)
Region: Podkarpackie		−0.031	−0.011
Region: Podlaskie		−0.150***	−0.148***
Region: Pomorskie		−0.006	−0.001
Region: Śląskie		−0.018	−0.000
Region: Świętokrzyskie		−0.129***	−0.111***
Region: Warmińsko-mazurskie		−0.026	−0.021
Region: Wielkopolskie		−0.024	−0.009
Region: Zachodniopomorskie		−0.042**	−0.037*
Survey year: 2012		Ref.	Ref.
Survey year: 2013		0.025**	0.017*
Survey year: 2014		−0.020**	−0.010
Constant	2.400***	1.456***	1.197***
Number of observations	14,118	14,118	14,118
R^2	0.042	0.191	0.212

One might be surprised at a negative wage premium from fluency in spoken and written Polish. This may be due to the fact that teachers and public servants, who have excellent Polish language skills, earn relatively low wages in Poland

***/**/* stand for 1%, 5%, and 10% significance, respectively

Table 10 Heckit estimates of the wage equation: full set of estimates. *Source* Author's own analyses based on Human Capital Balance survey data

Model specification	(1)	(2)	(3)
Elementary FLS	− 0.012	− 0.014	− 0.015
Intermediate FLS	0.130***	0.070***	0.052***
Advanced FLS	0.286***	0.153***	0.117***
Other skills (S_i):	No	No	Yes
Searching and analysing information and drawing conclusions			0.033***
Operating, installing and repairing technical equipment			0.002
Performing calculations			0.002
Operating a computer and using the Internet			0.041***
Artistic and creative abilities			0.000
Physical fitness			− 0.003
Self-organisation of work and showing initiative			0.000
Networking			− 0.012**
Arranging and performing office tasks			− 0.009*
Managerial skills and organising other people's work			0.038***
Availability			− 0.006
Fluency in spoken and written Polish			− 0.009*
Other control variables ($\boldsymbol{X}_i$):	No	Yes	Yes
Woman		− 0.186***	− 0.175***
Age		0.030***	0.027***
Age squared		− 0.000***	− 0.000***
Education: lower secondary or below		Ref.	Ref.
Education: basic vocational		0.044**	0.025
Education: secondary		0.158***	0.091***
Education: tertiary		0.429***	0.320***
Place of residence: rural		Ref.	Ref.
Place of residence: town < 10,000 inhabitants		0.015	0.006
Place of residence: town 10,000-19,999 inhabitants		0.043***	0.038***
Place of residence: town 20,000-49,999 inhabitants		0.074***	0.060***
Place of residence: town 50,000-99,999 inhabitants		0.096***	0.084***
Place of residence: town 100,000-199,999 inhabitants		0.113***	0.090***
Place of residence: town 200,000-499,999 inhabitants		0.109***	0.097***
Place of residence: town ≥ 500,000 inhabitants		0.106***	0.092***
Place of residence: Warsaw		0.194***	0.179***
Region: Dolnośląskie		Ref.	Ref.
Region: Kujawsko-pomorskie		− 0.056**	− 0.049**
Region: Lubelskie		− 0.087***	− 0.074***
Region: Lubuskie		0.028	0.042**
Region: Łódzkie		− 0.078***	− 0.067***
Region: Małopolskie		0.022	0.027
Region: Mazowieckie		0.052**	0.041*
Region: Opolskie		0.053**	0.068***

Table 10 (continued)

Model specification	(1)	(2)	(3)
Region: Podkarpackie		− 0.022	− 0.006
Region: Podlaskie		− 0.146***	− 0.143***
Region: Pomorskie		− 0.012	− 0.010
Region: Śląskie		− 0.018	− 0.004
Region: Świętokrzyskie		− 0.117***	− 0.101***
Region: Warmińsko-mazurskie		− 0.017	− 0.013
Region: Wielkopolskie		− 0.022	− 0.008
Region: Zachodniopomorskie		− 0.040**	− 0.037*
Survey year: 2012		Ref.	Ref.
Survey year: 2013		0.018*	0.010
Survey year: 2014		− 0.001	− 0.015
Constant	2.594***	1.699***	1.552***
Selection equation			
Elementary FLS	0.098***	0.099***	0.047
Intermediate FLS	0.099***	0.095***	− 0.024
Advanced FLS	0.011	− 0.035	− 0.215***
Searching and analysing information and drawing conclusions			0.024**
Operating, installing and repairing technical equipment			0.091***
Performing calculations			0.009
Operating a computer and using the Internet			0.080***
Artistic and creative abilities			− 0.056***
Physical fitness			0.123***
Self-organisation of work and showing initiative			0.100***
Networking			0.041***
Arranging and performing office tasks			0.019**
Managerial skills and organising other people's work			0.005
Availability			0.119***
Fluency in spoken and written Polish			− 0.081***
Age		0.284***	0.290***
Age squared		− 0.003***	− 0.003***
Education: lower secondary or below		Ref.	Ref.
Education: basic vocational		0.520***	0.425***
Education: secondary		0.748***	0.543***
Education: tertiary		1.227***	0.928***
Place of residence: rural		Ref.	Ref.
Place of residence: town < 10,000 inhabitants		0.025	− 0.024
Place of residence: town 10,000–19,999 inhabitants		0.018	− 0.002
Place of residence: town 20,000–49,999 inhabitants		0.081***	0.041
Place of residence: town 50,000–99,999 inhabitants		− 0.009	− 0.034
Place of residence: town 100,000–199,999 inhabitants		0.153***	0.103***
Place of residence: town 200,000–499,999 inhabitants		0.112***	0.060**
Place of residence: town ≥ 500,000 inhabitants		0.198***	0.186***

Table 10 (continued)

Model specification	(1)	(2)	(3)
Place of residence: Warsaw		0.730***	0.693***
Region: Dolnośląskie		Ref.	Ref.
Region: Kujawsko-pomorskie		− 0.264***	− 0.209***
Region: Lubelskie		− 0.385***	− 0.320***
Region: Lubuskie		− 0.127***	− 0.053
Region: Łódzkie		− 0.139***	− 0.092**
Region: Małopolskie		− 0.246***	− 0.204***
Region: Mazowieckie		− 0.085*	− 0.044
Region: Opolskie		− 0.265***	− 0.196***
Region: Podkarpackie		− 0.270***	− 0.157***
Region: Podlaskie		− 0.154***	− 0.163***
Region: Pomorskie		0.175***	0.246***
Region: Śląskie		− 0.141***	− 0.059
Region: Świętokrzyskie		− 0.384***	− 0.328***
Region: Warmińsko-mazurskie		− 0.264***	− 0.239***
Region: Wielkopolskie		− 0.093**	− 0.030
Region: Zachodniopomorskie		− 0.101**	− 0.007
Survey year: 2012		Ref.	Ref.
Survey year: 2013		0.143***	0.140***
Survey year: 2014		0.162***	0.139***
Woman	− 0.039**	− 0.246***	− 0.079***
Having a child aged 0–4 years	0.926***	0.506***	0.468***
Woman having a child aged 0–4 years	− 1.195***	− 1.016***	− 0.907***
Constant	− 0.306***	− 6.034***	− 7.805***
Number of observations: selection equation	35,563	35,563	35,563
Number of observations: wage equation	14,110	14,110	14,110
chi2	98.22	10.77	13.34
Prob > chi2	0	0.001	0.0002

The selection equation includes two variables: having a child aged 0–4 years, woman having a child aged 0–4 years; ***/**/* stand for 1%, 5%, and 10% significance, respectively

References

Adesope OO, Lavin T, Thompson T, Ungerleider C (2010) A systematic review and meta-analysis of the cognitive correlates of bilingualism. Rev Educ Res 80:207–245. https://doi.org/10.3102/00346543103688803

Albouy D (2008) The wage gap between Francophones and Anglophones: a Canadian perspective, 1970–2000. Can J Econ/Revue Canadienne d'Economique 41(4):1211–1238

Aldashev A, Gernandt J, Thomsen SL (2009) Language usage, participation, employment and earnings: evidence for foreigners in West Germany with multiple sources of selection. Labour Econ 16:330–341

Antonietti R, Loi M (2014) The demand for foreign languages in Italian manufacturing. AlmaLaurea working papers, no. 57

Azam M, Chin A, Prakash N (2013) The returns to English-language skills in India. Econ Dev Cult Change 61:335–367. https://doi.org/10.1086/668277

Berman E, Lang K, Siniver E (2003) Language skill complementarity: returns to immigrant language acquisition. Labour Econ 10(3):265–290
Bleakley H, Chin A (2010) Age at arrival, English proficiency, and social assimilation among US immigrants. Am Econ J Appl Econ 2(1):165–192
Casale D, Posel D (2011) English language proficiency and earnings in a developing country: the case of South Africa. J Soc Econ 40:385–393. https://doi.org/10.1016/j.socec.2011.04.009
Cattaneo A, Winkelmann R (2005) Earning differentials between German and French speakers in Switzerland. Swiss J Econ Stat 141(2):191–212
Chen Z, Lu M, Xu L (2014) Returns to dialect: identity exposure through language in the Chinese labor market. China Econ Rev 30:27–43
Chiswick BR, Miller PW (1995) The endogeneity between language and earnings: international analyses. J Labor Econ 13:246–288
Chiswick BR, Miller PW (2009) Earnings and occupational attainment among immigrants. Ind Relat 48(3):454–465
Chiswick BR, Miller PW (2010) Occupational language requirements and the value of English in the US labor market. J Popul Econ 23:353–372. https://doi.org/10.1007/s00148-008-0230-7
CSO (2013) Education in 2012/2013 School Year, Statistical Information and Elaborations. Central Statistical Office, Warsaw
Di Paolo A (2011) Knowledge of Catalan, public/private sector choice and earnings: evidence from a double sample selection model. Hacienda Pública Española/Rev Publ Econ 197(2):9–35
Di Paolo A, Raymond JL (2012) Language knowledge and earnings in Catalonia. J Appl Econ 15(1):89–118
Di Paolo A, Tansel A (2015) Returns to foreign language skills in a developing country: the case of Turkey. J Dev Stud 51(4):407–421
Dustmann C (1994) Speaking fluency, writing fluency and earnings of migrants. J Popul Econ 7(1):133–156
Dustmann C, Fabbri F (2003) Language proficiency and labour market performance of immigrants in the UK. Econ J 113(489):695–717
Dustmann C, van Soest A (2001) Language fluency and earnings: estimation with misclassified language indicators. Rev Econ Stat 83(4):663–674
Dustmann C, van Soest A (2002) Language and the earnings of immigrants. Ind Labor Relat Rev 55(3):473–492
Fabo B, Beblavy M, Lenaerts K (2017) The importance of foreign language skills in the labour markets of Central and Eastern Europe: assessment based on data from online job portals. Empirica 44:487–508. https://doi.org/10.1007/s10663-017-9374-6
Fry R, Lowell L (2003) The value of bilingualism in the U.S. labor market. Ind Labor Relat Rev 57(1):128–140
Ginsburgh VA, Prieto-Rodriguez J (2011) Returns to foreign languages of native workers in the EU. Ind Labor Relat Rev 64(3):599–617
Ginsburgh VA, Prieto-Rodriguez J (2013) Is there a gender bias in the use of foreign languages in Europe? Kyklos 66(4):552–566
Grin F (2001) English as economic value: facts and fallacies. World Englishes 20(1):65–78
Grin F, Sfreddo C (1998) Language-based earnings differentials on the Swiss labour market: is Italian a liability? Int J Manpow 19(7):520–532
Guo Q, Sun W (2014) Economic returns to English proficiency for college graduates in mainland China. China Econ Rev 30:290–300
Heckman J (1979) Sample selection bias as a specification error. Econometrica 47(1):153–161
Henley A, Jones RE (2005) Earnings and linguistic proficiency in a bilingual economy. Manch School 73(3):300–320
Isphording I (2015) Language and labor market success. International encyclopedia of the social and behavioral sciences, 2nd edn. Elsiever, Amsterdam, pp 260–265
Ku H, Zussmann A (2010) Lingua Franca: the role of English in international trade. J Econ Behav Organ 75:250–260. https://doi.org/10.1016/j.jebo.2010.03.013
Lang K, Siniver E (2009) The return to English in a non-English speaking country: Russian immigrants and native Israelis in Israel. BE J Econ Anal Policy 9:1–30. https://doi.org/10.2202/1935-1682.2274
Leslie D, Lindley J (2001) The impact of language ability on employment and earnings of Britain's ethnic communities. Economica 68(272):587–606

Levinsohn J (2007) Globalization and the returns to speaking English in South Africa. In: Harrison AE (ed) Globalization and poverty. University of Chicago Press, Chicago, pp 629–646
Rendon S (2007) The Catalan premium: language and employment in Catalonia. J Popul Econ 20(3):669–686
Rubin DB (2001) Using propensity scores to help design observational studies: application to the tobacco litigation. Health Serv Outcomes Res Method 2(3):169–188
Saiz A, Zoido E (2005) Listening to what the world says: bilingualism and earnings in the United States. Rev Econ Stat 87(3):523–538
Shapiro D, Stelcner M (1997) Language earnings in Quebec: trends over twenty years 1970–1990. Can Public Policy 23(2):115–140
Shields MA, Price WS (2002) The English language fluency and occupational success of ethnic minority immigrant men living in English metropolitan areas. J Popul Econ 15(1):137–160
Spence AM (1973) Job market signaling. Q J Econ 87(3):355–374
Stöhr T (2015) The returns to occupational foreign language use: evidence from Germany. Labour Econ 32:86–98
Toomet O (2011) Learn English, not the local language! Ethnic Russians in the Baltic States. Am Econ Rev Pap Proc 101:526–531. https://doi.org/10.1257/aer.101.3.526
Wang H, Smyth R, Cheng Z (2017) The economic returns to proficiency in English in China. China Econ Rev 43:91–104. https://doi.org/10.1016/j.chieco.2017.01.004
Williams D (2011) The economic returns to multiple language usage in Western Europe. Int J Manpow 32:372–393. https://doi.org/10.1108/01437721111148513