Educational mismatch in the Czech Labour Market

Martina Mysíková[1]

Abstract: Educational mismatch in labour markets is a phenomenon that has been widely analysed, mainly with respect to rising concerns about a possible oversupply of graduates. Like most European countries, the Czech Republic has experienced a boom in tertiary education in the last decade. The incidence and determinants of over- and undereducation vary substantially depending both on the mismatch measurement approach and the data source applied. Educational mismatch is also reflected in wage levels: overeducated workers have lower wages and undereducated workers have higher wages than workers with the same education whose jobs match their education level. Second, overeducated workers earn more and undereducated workers earn less than their co-workers with exactly the required level of education. The effects are qualitatively the same regardless of the data source and measurement approach applied, but their sizes differ slightly.

Key words: educational economics, human capital, overeducation, undereducation, wages

JEL Classification: I21, J24, J31

Received: 2 October 2015 / Accepted: 27 January 2016/ Sent for Publication: 16 June 2016

Introduction

Educational mismatch in labour markets is a phenomenon that has recently received a lot of attention, mainly due to a growing number of university graduates and related concerns about a possible oversupply of graduates. Many studies have looked at educational mismatch on the U.S. labour market (among others, McGoldrick and Robst, 1996, analysed overeducation probability; Rubb, 2003, discussed the persistence of overeducation). In the European context, most studies are devoted to Western European countries (Bauer, 2002, analysed the wage effects of educational mismatch in Germany; Karakaya, Plasman and Rycx, 2007, examined the determinants of overeducation in Belgium; Kiker, Santos and De Oliveira, 1997, looked at the determinants and wage effects of overeducation in Portugal); far fewer have focused on Eastern Europe.

Education in the Czech Republic has changed since the beginning of the 1990s. First, as in other post-communist countries, new graduates shifted away from technical fields towards business fields during the transition from a planned to a market-based economy

---

[1] National Training Fund, Opletalova 1, 110 00 Praha 1

(Jeong, Kejak and Vinogradov, 2008). Second, as in most European countries, the Czech Republic has experienced a boom in tertiary education in the last decade, resulting in the threat of an oversupply of graduates, and increasing the chances of overeducation (or even unemployment). So far, these issues have not been thoroughly addressed in the Czech Republic.

An excess supply of individuals with higher education is believed to result in them occupying jobs for which they are overqualified. As a consequence, workers with lower education are pushed out of positions suited to their education level into less demanding jobs. As the number of people receiving higher education further increases, this problem could intensify. According to Freeman (1976), if an excess of qualified workers are pushed into 'lower-qualified' jobs, returns to education should fall, which would lower investments in education and as a consequence the labour market would ultimately reach equilibrium. Chevalier (2003) argues that this did not happen in the U.S. because returns to education remained high. Gottschalk and Hansen (2003) did not find any evidence to support the idea that an increasing share of college graduates had been forced into non-college jobs between the mid-1980s and the mid-1990s in the U.S.

Similarly, studies on education levels in the UK have argued that although the share of tertiary educated graduates has increased over the last few decades, returns to education have remained stable (Chevalier, 2003; Bevan and Cowling, 2007). Contrary to expectations, the growing share of the more highly educated together with stable returns from education indicate that demand for graduate workers has coped well with the increasing supply. Some researchers have looked for alternative explanations of the fact that returns to education have remained stable, for instance, skill heterogeneity among workers with the same education level (for instance, Chevalier, 2003).

While much has been written on these issues in advanced economies, information relevant to transitional economies is rather scarce. This study aims to describe the situation in the Czech Republic (hereafter CR), where returns to education have risen rapidly since the end of the communist era in 1989, reflecting changed principles of remuneration. According to Večerník (2013), returns to education doubled during the early years of economic transition but have remained stable since 1996 for both men and women. This suggests that since the late 1990s the 'Western' patterns of stable returns to education alongside rising tertiary enrolments may have applied in the CR as well. According to the Eurostat database, the share of the working-age population with tertiary education increased from 10 percent in 2003 to 17 percent in 2012. This growth mainly occurred among young people: in 2003, 13 percent of 30-34 year-olds had tertiary education; this share had risen to 26 percent by 2012.

As there is little existing empirical evidence of educational mismatch in the Czech labour market, and in order to view this phenomenon as broadly as possible, this study applies various methodological approaches to measure the level of education required by different occupations, as used in the relevant literature. Measurements of educational mismatch are highly sensitive to the methodology used, as I describe in the next section. For that reason, this study uses several data sources to check the robustness of the results. The third section describes the EU-SILC and PIAAC data used, and comments on the potential pitfalls of these datasets. The fourth section analyses determinants of over-

and undereducation, while the fifth section measures the effects of educational mismatch on wages. The final section evaluates the results obtained under different methodological approaches and data sources.

Approaches to measuring required education levels

Analyses of educational mismatch typically face one problem: the results tend to differ according to the methodology used to measure the education level required in a certain occupation. Workers are considered adequately educated (matched), if their education fits (but does not exceed) the requirements of their job. In all other cases, they are considered overeducated/undereducated: their level of education either exceeds or falls short of the requirements of their job.

Three approaches to determining the education level required for a given occupation appear in the literature. OECD (2007, p. 135) calls these approaches normative, statistical, and self-declared. The *normative* approach relates education and job qualifications based on an exogenous definition of education requirements determined by job analysts (used, for example, by Chevalier, 2003 and Karakaya et al., 2007). In European practice, OECD methodology (OECD, 2007) is usually applied; this links educational ISCED codes with ISCO classifications according to an established scheme (see for instance Hernández and Serrano, 2012). Studies on the U.S. labour market typically use the Dictionary of Occupational Titles for this purpose (e.g. McGoldrick and Robst, 1996).

The *statistical* approach determines the required number of years of education for a particular job by observing realized job matches. Two measures under this approach have been discussed in the literature: these express the number of years of education required either as a) a one-standard-deviation range around the national mean number of years of education in a given occupation (Verdugo and Verdugo, 1989; Bauer, 2002; Rubb, 2003) or b) a modal value (suggested by Kiker et al., 1997 and applied, for instance, by Bauer, 2002 and Karakaya et al., 2007).

The third, *self-declared approach* uses workers' answers to questions about the requirements of their jobs (Nieto, 2014; Alba-Ramírez, 1993; McGoldrick and Robst, 1996; Rašovec and Vavřinová, 2014).

McGoldrick and Robst (1996) used all three approaches to analyse the situation in the U.S. in 1985. Following the normative approach they classified more than half of both men and women as overeducated. On the other hand, according to the statistical approach (one-standard-deviation measure), only 16 percent of male and 9 percent of female workers were overeducated, while the self-declared approach reported about 30 percent of men and women overeducated. Rašovec and Vavřinová (2014) used direct and indirect self-declared approaches to measure educational mismatch in the Czech Republic and found that the share of undereducated individuals in particular differed substantially depending on the exact approach used. Chevalier (2003) notes that the choice of approach highly influences the incidence of overeducation, but not its effects on wages.

Another problem is that all three approaches assume that individuals with the same level of education are homogeneous. Some studies control for unobserved heterogeneity

using panel estimation techniques (for instance, Bauer, 2002). Chevalier (2003) deals with the heterogeneity of UK graduates by creating a group of 'clever ones' and a group of 'under-achievers', based on a question about how dis/satisfied the respondents are with the match between their work and their qualifications (assuming that overeducated individuals who are satisfied are only 'apparently' overeducated, while those who are dissatisfied are 'genuinely' overeducated). Unfortunately, none of the available Czech data sources seem to provide any means by which the homogeneity assumption might be adequately examined.

All three approaches to measuring required education levels are also subject to other criticisms. The normative approach is often criticised for its arbitrary nature, especially if the same established scheme is applied across different countries, since the level of education required for a given occupation can differ from one country to another (e.g. OECD, 2007, p.135). The statistical approach has been criticised for its arbitrariness, as there is no logic behind the choice of one standard deviation (McGoldrick and Robst, 1996). On the one hand, this approach has the advantage of being sensitive to technological changes and labour market characteristics (Karakaya et al., 2007), but on the other hand, the reference group of matched workers could itself in fact be overqualified if the majority of workers in the given occupation are more educated than actually needed. The advantage of the self-declared approach is that it is job-specific; its disadvantage is its subjectivity and respondents' tendency to overestimate their own qualification for the job they currently perform.

Data for the Czech Republic and applicability of existing approaches

I use three data sources for my analyses. The first is the national Labour Force Survey (LFS) from the first quarter of 2013, with a total sample of 57,000 respondents; this is the biggest sample appropriate individual/household survey in the CR. However, the LFS does not collect information on wages and thus cannot be used for the last part of my analysis. Second, I apply the national version of the Statistics on Income and Living Conditions (SILC) survey from 2012, with a total sample of almost 21,000 respondents. The national dataset includes the variable of educational attainment in a more detailed structure than the international dataset (for more details on the differences between the Czech national and international SILC datasets see Mysíková, 2011). Third, I apply data from the national survey of the Programme for the International Assessment of Adult Competencies (PIAAC), conducted in the CR in 2011/2012, with a total sample of 6,000 respondents.

In all three datasets, I select working age (16-65) full-time employees working 30+ hours per week. For the normative approach, I follow the OECD methodology for measuring educational mismatch, according to which low-skilled workers with ISCED level 0-2 education are matched in low-skilled occupations with ISCO score 9, and undereducated otherwise. Intermediate-skilled workers with ISCED level 3-4 education are matched at intermediate occupations with ISCO score 4-8, overeducated if they work in ISCO score 9 occupations and undereducated if they work in ISCO score 1-3 occupations. Finally, highly-skilled workers with ISCED level 5-6 education are

matched in high-skilled occupations with ISCO score 1-3 and overeducated otherwise (for the detailed scheme, see OECD, 2007, p. 156).

Educational attainment was rescaled into 9 categories (ISCED 1, 2, 3a-b, 3c, 4, 5a – bachelor, 5a - master, 5b, 6), consistent throughout the datasets. For the statistical approach, years of education were derived based on the national standard length of time required to complete each education level. Due to it large sample size, the LFS dataset was used to determine the one-standard-deviation range around the average years of education in each occupation, and these required years of education were applied in SILC and PIAAC.[2] The statistical approach is based on 1-digit ISCO scores. This simplification does not alter the results; when applied on LFS data, the statistical approach had almost the same results regardless of the level of detail of ISCO classification used.[3]

The self-declared approach can only be applied using the PIAAC dataset. I used the question: 'If applying today, what level of education would a potential candidate need in order to get this job?' The possible answers correspond to the ISCED codes. The education levels stated by the respondents were considered to represent the education level required in the respondents' jobs.

Table 1 shows that the normative approach yields the lowest share of overeducation (6.0-8.3 percent) and the highest share of undereducation (21.9-26.1 percent); this suggests that undereducation is a far more frequent phenomenon than overeducation. The results are quite different when the other approaches are applied. With the statistical approach, the share of overeducation was about 10 percentage points higher (14.2-18.5 percent), yet the share of undereducation was the lowest, especially when using LFS and SILC data (12.4-12.7 percent). The self-declared approach showed a striking one quarter of respondents to be overeducated, and the share of matched respondents was by far the lowest.

---

[2] When SILC or PIAAC means and standard deviation values were used to determine the required years of education in an occupation, the share of overeducation was by 1.8 pp higher and the share of undereducation by 1.8 pp lower in SILC. In PIAAC, the share of overeducation was by 0.5 pp lower and the share of undereducation by 2.8 pp lower, compared to using the LFS values in particular data.

The "mode" measure of the statistical approach was examined as well. It resulted in 5-8 pp higher share of overeducation and 2-6 pp higher share of undereducation. The matched group of workers is defined more strictly by the "mode" than the "mean" measure. The one-standard-deviation range around the mean of derived years of education usually does not include only one education level but also adjacent levels, as against the modal value which determines just one educational level which is the required one in an occupation.

[3] The broader the ISCO categories, the higher were the shares of both over- and undereducated; however, the difference was only moderate: The share of overeducated ranged from 14.2% to 14.5% and the share of undereducated ranged from 10.0% to 12.4% resulting from 3-digit and 1-digit ISCO, respectively.

Table 1 Educational mismatch (%)

| | Normative | | | Statistical | | | Self-declared |
	LFS	SILC	PIAAC	LFS	SILC	PIAAC	PIAAC
Total							
Overeducated	6.6	6.0	8.3	14.5	14.2	18.5	25.4
Matched	71.6	67.9	69.5	73.0	73.2	64.9	59.8
Undereducated	21.9	26.1	22.1	12.4	12.7	16.7	14.9
Male							
Overeducated	5.0	4.9	5.9	17.5	17.2	23.8	25.4
Matched	74.8	73.7	75.0	72.1	71.3	62.8	61.5
Undereducated	20.1	21.4	19.1	10.3	11.4	13.4	13.1
Female							
Overeducated	8.6	7.2	11.3	10.7	10.4	21.3	25.3
Matched	67.4	60.9	63.0	74.1	75.4	55.7	57.7
Undereducated	24.1	31.8	25.7	15.2	14.2	23.0	17.0
Youth (16-24)							
Overeducated	7.5	6.8	8.0[a]	21.0	15.3	20.5[a]	25.0[a]
Matched	72.7	70.1	77.0[a]	68.0	73.2	65.2[a]	64.8[a]
Undereducated	19.8	23.0	15.0[a]	11.0	11.4	14.3[a]	10.2[a]
Elderly (55-65)							
Overeducated	5.9	6.6	4.9[a]	10.2	12.3	21.9[a]	22.2[a]
Matched	69.0	64.0	66.8[a]	71.9	69.2	50.1[a]	53.4[a]
Undereducated	25.1	29.3	28.3[a]	17.9	18.5	28.0[a]	24.4[a]

Source: National LFS 1q2013, national SILC (Životní podmínky) 2011, PIAAC. Author's compu-tations.
Notes: Unweighted sample size: LFS – 18,647; SILC – 7,044; PIAAC – 2,344. [a]Sample of youth and elderly in PIAAC includes fewer than 340 observations.

While all the approaches concur that undereducation is higher for women than for men, the normative approach suggests that women suffer from greater overeducation than men, while the statistical approach suggests that men are more often overeducated than women. With the exception of the PIAAC data, the lower sections of Table 1 show that the incidence of overeducation decreases with age. All the data sources and approaches confirm that the incidence of undereducation increases with age.
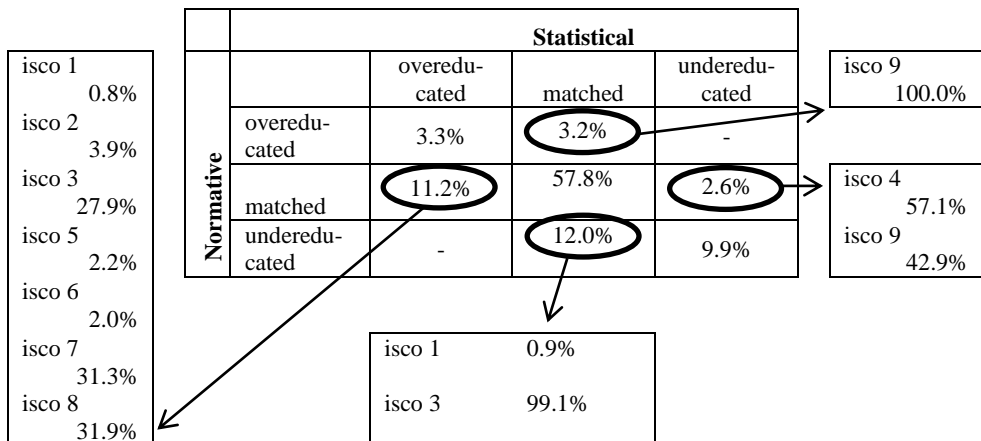
There are several clear reasons for the differences between the normative and the statis-tical approaches. First, workers with ISCED level 6 education (PhD degrees) are always classified as overeducated under the statistical approach, while under the normative approach they are considered matched if they work in ISCO score 1-3 occupations. At the other end of the spectrum, workers with ISCED level 1-2 education (primary school) are always classified as undereducated by the statistical approach, although under the

normative approach they are considered matched if they work in ISCO score 9 occupa-tions. These facts demonstrate the disadvantages of the statistical approach, as it is clear that in reality both workers with PhDs and workers with primary education could theo-retically find jobs whose requirements fit their education level. Nevertheless, since only 0.6 percent of my LFS sample are PhD holders and 4.2 percent are primary-educated, the bias caused by the statistical approach in this particular study may be negligible.

Figure 1 depicts the differences between the normative approach and the mean measure of the statistical approach. In total, the normative approach imposes higher educational requirements than the statistical approach on 23.2 percent of workers. The normative approach classifies 11.2 percent of workers as matched, while these are considered overeducated under the statistical approach. These are mainly in ISCO score 3, 7 and 8 occupations and are workers with ISCED level 3a-b education (63.9 percent) or ISCED level 5a– master degree education (27.6 percent). Further, 12.0 percent of workers are classified as undereducated under the normative approach while their education and occupation are matched under the statistical approach. These are mainly ISCO score 3 occupations and workers with ISCED level 3a-b education.

Meanwhile, the statistical approach imposes higher educational requirements than the normative approach on 5.8 percent of workers. 3.2 percent of workers are matched according to the statistical approach but classified as overeducated according to the normative approach. These employees work solely in ISCO 9 occupations and their education level is ISCED 3c. Finally, 2.6 percent of workers are undereducated accord-ing to the statistical approach but matched according to the normative approach. These employees work solely in ISCO 4 or 9 occupations and their education level is either ISCED 2 (42.9 percent) or ISCED 3c (57.1 percent).

Figure 1 Differences in educational mismatch under normative and statistical approaches using LFS data

		Statistical		
		overedu-cated	matched	underedu-cated
Normative	overedu-cated	3.3%	3.2%	-
	matched	11.2%	57.8%	2.6%
	underedu-cated	-	12.0%	9.9%

isco 1   0.8%
isco 2   3.9%
isco 3   27.9%
isco 5   2.2%
isco 6   2.0%
isco 7   31.3%
isco 8   31.9%

isco 9   100.0%

isco 4   57.1%
isco 9   42.9%

isco 1   0.9%
isco 3   99.1%

Source: National LFS 1q2013. Author's computations.

The normative approach based on the OECD scheme may not be suitable for the CR, given that it reports relatively low overeducation but relatively high undereducation – this is the opposite to what we would expect given the current labour market and increasing levels of tertiary education. One third of Czech workers with ISCED level 3a-b education are employed in jobs with ISCO score 3, for which the normative approach considers them undereducated.

As expected following similar experience in other countries (see e.g. McGoldrick and Robst, 1996, for the U.S.), the outcome of educational mismatch analysis for the CR is to a large extent dependent on the measurement approach applied. However, it is clear that the data source also influences the results; I identify a difference of up to 4 percentage points in the levels of both over- and undereducation when using the same approach but a different data source. The next section examines the determinants of over- and undereducation and demonstrates the dissimilarities between the different approaches and different datasets.

Determinants of educational mismatch

Tables 2 and 3 show the impacts of various characteristics on the likelihood of being overeducated or undereducated under the three approaches, based on the SILC and PIAAC datasets. A multivariate logit regression was applied and hence the estimated logit coefficients must be interpreted relative to the reference group, which is matched respondents. The regressions were estimated separately for men and women when SILC data was used, but not when using data from the much smaller PIAAC survey.

Under the normative approach (Table 2), men are overall less likely than women to be over- or undereducated in their jobs. The self-declared approach (Table 3) produces the same results regarding undereducation (but statistically insignificant regarding overeducation). However, the statistical approach produces the opposite result for overeducation: the probability of being overeducated rather than matched is higher for men than for women. Theoretically, women would be expected to be overeducated more often than men, for several reasons such as vertical segregation, gender differences in time constraints and geographical mobility (as secondary earners, women more often move to partners' job locations and are more likely to be constrained in their job search by family responsibilities). However, my results from the statistical approach contradict these expectations. The literature offers examples of women having both higher (for instance, Karakaya et al., 2007, for Belgium) and lower (for example, Kiker et al., 1997, for Portugal) probability of being overeducated for their jobs than men.

The more years of education an individual has, the higher the probability that they are overeducated: ,my results mostly confirm this positive relation (similarly to Kiker et al., 1997, for Portugal), with the only exception being the normative approach based on SILC data, where the negative impact is given by the female subsample.

Table 2 Determinants of overeducation and undereducation: Normative approach

| | SILC | | | | | | PIAAC | |
| | Total | | Male | | Female | | Total | |
	Over	Under	Over	Under	Over	Under	Over	Under
Years of education	-0.037	-0.327***	0.009	-0.367***	-0.084**	-0.343***	0.133***	-0.326***
Male	-0.415***	-0.620***	-	-	-	-	-0.699***	-0.713***
Experience	-0.045**	-0.006	-0.076***	-0.033**	-0.012	0.014	0.001	0.011
Experience²	0.001***	0.000	0.002***	0.001**	0.000	-0.000	-0.000	-0.000
Prague	0.012	0.653***	0.052	0.641***	-0.078	0.577***	-0.430	-0.084
Supervisory	-0.688***	1.274***	-0.286	1.800***	-1.569***	0.778***	-0.787***	1.322***
Unlimited contract	-0.650***	0.216*	-0.268	-0.036	-0.882***	0.422**	-0.524**	0.399**
Firm size 50+	0.083	0.082	0.161	0.193*	-0.040	-0.115	0.733***	0.065
Constant	-0.876**	2.334***	-2.052***	2.654***	-0.304	2.075***	-2.225***	1.297**
Pseudo R²	0.10		0.10		0.11		0.12	
LR X² (34/32)	1004.2		506.85		562.87		387.53	

Source: National SILC (Životní podmínky) 2011, PIAAC. Author's computations.
Notes: Weighted. *** statistically significant at the 1% level, ** statistically significant at the 5% level, * statistically significant at the 10% level. All regressions include nine dummies for industry.

111

Table 3 Determinants of overeducation and undereducation: Statistical and self-declared approach

		Statistical								Self-declared	
		SILC						PIAAC		PIAAC	
		Total		Male		Female		Total		Total	
		Over	Under	Over	Under	Over	Under	Over	Under	Over	Under
Years of education		0.540***	-1.144***	0.466***	-1.140***	0.749***	-1.253***	0.622***	-0.965***	0.278***	-0.397***
Male		0.654***	-0.477***	-	-	-	-	1.039***	-0.939***	-0.107	-0.460***
Experience		0.009	-0.039**	0.026	-0.060***	0.015	-0.020	-0.049*	0.008	0.000	-0.017
Experience^2		-0.000	0.001***	-0.001	0.001***	-0.001	0.001	0.001	0.000	0.000	0.001*
Prague		-0.628***	0.436***	-0.310*	0.702***	-1.232***	-0.019	-0.468*	0.082	-0.309*	-0.231
Supervisory		-1.250***	2.203***	-1.252***	2.544***	-1.070***	1.921***	-0.750***	2.432***	-0.873***	1.145***
Unlimited contract		-0.130	0.270	0.002	-0.128	-0.371	0.722***	0.088	-0.016	-0.290*	-0.046
Firm size 50+		-0.328***	0.183*	-0.174	0.073	-0.724***	0.251	-0.029	-0.426***	0.009	0.129
Constant		-8.958***	10.828***	-7.835***	11.096***	-11.282***	11.055***	-10.320***	9.050***	-3.626***	2.532***
Pseudo R^2		0.26		0.22		0.35		0.32		0.11	
LR X^2 (34/32)		2555.55		1226.42		1468.95		1138.81		429.23	

Source: National SILC (Životní podmínky) 2011, PIAAC. Author's computations.
Notes: Weighted. *** statistically significant at the 1% level, ** statistically significant at the 5% level, * statistically significant at the 10% level. All regressions include nine dummies for industry.

Many existing studies discuss a possible trade-off between education and other forms of human capital, such as experience or on-the-job training, whereby workers may substitute their lack of work experience with overeducation and accept a job that requires a lower education than they have achieved in order to gain more experience, which will later enable them to improve their job level (Rosen, 1972). In this case, overeducation would be a temporary phenomenon.

On the other hand, insufficient years of education might be compensated by experience in some cases, which would result in undereducation being long-lasting or permanent. In my analysis, the negative relationship between experience and the probability of being overeducated proved to be statistically significant under the normative approach based on SILC data (for men only) and under the statistical approach based on PIAAC data. However, the positive relationship between experience and the probability of being undereducated did not prove significant in any case; on the contrary the opposite holds for men based on the SILC data.

Living in the capital city decreases the probability of overeducation compared to having a matching job, although no such significant effect is identified under the normative approach. Meanwhile, living in Prague increases the likelihood of being undereducated rather than matched when SILC data are used. These results demonstrate that the larger labour market in the capital city provides more vacancies and workers thus have a better chance of finding a job that matches their education level rather to be overeducated in "worse" jobs, as well as they have more opportunities to apply even for jobs with higher educational requirements. No such effect was observed with the PIAAC data.

Supervisory positions are less likely than non-supervisory positions to be occupied by overeducated individuals and more likely to be occupied by undereducated individuals (relative to matched workers); this is uniform across the approaches and datasets. These findings seem surprising, especially where undereducation is concerned, since it would seem logical to expect supervisory positions to be held by adequately educated or overeducated workers. Note, however, that when an overeducated worker is promoted to a position demanding a higher educational level s/he becomes matched to his/her job. In the same way, if a matched worker is promoted to a position that demands a higher educational level, for instance due to her/his other qualifications and skills, then s/he becomes undereducated. This could explain why the share of undereducated individuals is higher among supervisory positions than among non-supervisory positions, and this suggests that skills and qualification characteristics other than an individual's formal education level play a role in promotion.

All approaches except the statistical one confirm that having a permanent job contract (compared to all other forms of contracts) usually decreases the probability of being overeducated (rather than matched) and increases the probability of being undereducated. The models based on SILC data were run separately for men and women and reveal that this in fact holds true only for women. No statistically significant impact was found for men. These findings are in contrast to those by Karakaya et al. (2007) for Belgium, where permanent job contracts were found to be positively related to the probability of being overeducated.

Van der Meer and Wielers (1996) found a positive relationship between firm size and overeducation. The explanation for this is that large firms have higher costs for monitoring productivity, and so they employ more educated workers, who are granted more autonomy. Karakaya et al. (2007) did not find this to be true when examining the Belgian labour market. My results do not provide any consistent findings in this regard. While company size does not have any statistically significant impact in most cases, workers in large firms (with 50 and more employees) are shown to be more likely to be overeducated (rather than matched) than those in smaller firms under the normative approach based on PIAAC (Table 2), and less likely under the statistical approach based on SILC data (Table 3).

Educational mismatch and wages

Two models are commonly applied in the literature to estimate the effects of overeducation and undereducation on wages: the first of these follows Verdugo and Verdugo (1989) and the second follows Duncan and Hoffman (1981). I apply both models, using all three approaches to measure educational mismatch and both the SILC and PIAAC datasets. While the results in the previous sections showed that the shares of over- and undereducation and the impacts of various determinants on the probability of being over- and undereducated differ according to the approach and data used, the results of the models applied in this section suggest that the associated wage effects are rather consistent.

The Verdugo and Verdugo model (1989) can be written as:

$$lnY_i = \alpha_0 + \alpha_1 E_i + \alpha_2 OVER_i + \alpha_3 UNDER_i + X_i\gamma + \varepsilon_i, \tag{1}$$

where lnY_i is the log of gross hourly wages for individual i, E_i denotes the individual's actual number of years of education attained, $OVER_i$ and $UNDER_i$ are dummies for overeducation and undereducation, respectively, X_i is a vector including other explanatory variables with the vector of coefficients γ, and ε_i is an error term.

The Verdugo and Verdugo model includes the number of years of education attained, and compares mismatched and matched workers with the same level of education. If productivity and wages were determined by actual years of education attained, then the education level required in a particular job and whether an individual is over- or undereducated in that job would be irrelevant to wage determination: the coefficients α_2 and α_3 would be zero. In other words, only actual years of education would be relevant in determining the individual's wage level. If $\alpha_2 = \alpha_3 = 0$, equation (1) would be reduced to a standard human capital Mincerian (1974) wage equation:

$$lnY_i = \alpha_0 + \alpha_1 E_i + X_i\gamma + \varepsilon_i. \tag{2}$$

If, on the other hand, wages are related to the number of years of education required by a particular job, any additional years of actual education would not be rewarded and would not increase the individual's wage. Hence an overeducated worker would earn less than a matched worker with the same actual education level: coefficient α_2 would be negative. In the same logic, an undereducated worker would earn more than an

equally educated worker in a job with lower educational requirements that actually fit his education level.

The last row of Table 4, which reports the results of equation (1) using an OLS regression, shows that based on computed F-statistics the null hypothesis $\alpha_2 = \alpha_3 = 0$ can be rejected regardless of the approach and data used.

Table 4 Wage models: Verdugo–Verdugo

	Normative				Statistical				Self-declared
	SILC			PIAAC	SILC			PIAAC	PIAAC
	Total	Male	Female	Total	Total	Male	Female	Total	Total
Years of educ.	0.071*	0.056*	0.088*	0.060*	0.075*	0.061*	0.095*	0.069*	0.064*
OVEReducated	-0.164*	-0.150*	-0.162*	-0.191*	-0.102*	-0.082*	-0.140*	-0.146*	-0.161*
UNDEReduc.	0.117*	0.093*	0.146*	0.137*	0.080*	0.078*	0.103*	0.104*	0.073*
R^2	0.43	0.35	0.48	0.44	0.42	0.35	0.45	0.42	0.43
H_0: $\alpha_2 = \alpha_3$	160.12*	48.10*	129.70*	467.96*	57.97*	25.18*	41.89*	37.09*	70.70*

Source: National SILC (Životní podmínky) 2011, PIAAC. Author's computations.
Notes: Weighted. Unweighted sample size: SILC – 6,471 (3,511 men, 2,960 women); PIAAC – 2,034 (and 2,022 in self-declared approach).
* statistically significant at the 1% level. All regressions include a dummy for sex, experience and experience squared, a dummy for the Prague region, a dummy for supervisory positions, a dummy for unlimited job contracts, a dummy for large company size and nine dummies for industries. Male and female wage equations include also a dummy for parenthood and a dummy for marriage (as these have an opposite effect on wages for both men and women).

Overeducated workers earn 10 percent (statistical approach with SILC) or 19 percent (normative approach with PIAAC) less than workers with the same education whose jobs match their education level. The normative approach suggests this wage penalty is higher compared with the statistical and self-declared approaches. As far as the data sources are concerned, PIAAC data reports a higher wage penalty than SILC data for both the normative and statistical approaches. The wage penalty is higher for women than for men. Bauer (2002) reported slightly higher results using German Socio-Economic Panel data and the statistical approach: he found the wage penalty for overeducated men to be 10.6 percent and for overeducated women 15.1 percent; the equivalent results in this study are 8.2 percent and 14.0 percent, respectively. As Kiker et al. (1997) show, the wage penalty was higher for men (9.5 percent) than for women (3.6 percent) in Portugal in 1991.

Undereducated workers receive between 7 percent (self-declared approach with PIAAC data) and 14 percent (normative approach with PIAAC data) more pay than workers with the same education level who are in matched jobs. This wage premium is highest when the normative approach is used and lowest with the self-declared approach. PIAAC data produces higher wage premium results than SILC data. The wage premium is

higher for women than for men. Using the statistical approach, Bauer (2002) obtained a similar wage premium for undereducated men (8 percent) but a much lower premium for undereducated women (3 percent) in Germany, while Kiker et al. (1997) identified a c. 16 percent wage premium for both undereducated men and women in Portugal.

The Duncan and Hoffman (1981) model distinguishes the number of years of education required in an occupation (E_i^R), the number of years of overeducation (E_i^O) and the number of years of undereducation (E_i^U) such that actual years of education (E_i) are defined by:

$$E_i = E_i^R + E_i^O - E_i^U. \tag{3}$$

For overeducated workers, $E_i^O > 0$ and $E_i^U = 0$, while for undereducated workers, $E_i^O = 0$ and $E_i^U > 0$. I apply this decomposition to the statistical and self-declared approaches only, since the normative approach groups several ISCED educational levels together and, hence, the number of years of education required is not specifically determined. The actual years of education then enter the Duncan and Hoffman's specification as follows:

$$lnY_i = \beta_0 + \beta_1 E_i^R + \beta_2 E_i^O + \beta_3 E_i^U + X_i\gamma + \varepsilon_i. \tag{4}$$

Coefficients β_2 and β_3 (returns to a year of education beyond and below the required, respectively) differ from those in the Verdugo and Verdugo model as here they are interpreted relative to co-workers who only have the required education level, in other words, relative to workers in the same occupation but with matched education.

Table 5 Wage models: Mincer and Duncan–Hoffman

	Statistical				Self-declared
	SILC			PIAAC	PIAAC
	Total	Male	Female	Total	Total
Years of education	0.064*	0.051*	0.080*	0.052*	0.052*
R^2	0.40	0.34	0.43	0.40	0.40
Years of education required	0.092*	0.072*	0.118*	0.095*	0.062*
Years of overeducation	0.040*	0.035*	0.047*	0.013[+]	0.025*
Years of undereducation	-0.051*	-0.035*	-0.063*	-0.056*	-0.030*
R^2	0.42	0.35	0.47	0.44	0.43
H$_0$: $\beta_1 = \beta_2 = -\beta_3$	101.20*	33.11*	92.43*	71.36*	73.46*
H$_0$: $\beta_2 = \beta_3 = 0$	202.11*	63.44*	153.69*	47.90*	26.62*

Source: National SILC (Životní podmínky) 2011, PIAAC. Author's computations.
Notes: Weighted. Unweighted sample size: SILC – 6,471 (3,511 men, 2,960 women); PIAAC – 2,034 (and 2,022 in self-declared approach).
* statistically significant at the 1% level. $^+$ statistically significant at the 5% level. All regressions include a dummy for sex, experience and experience squared, a dummy for Prague region, a dummy for supervisory position, a dummy for unlimited job contract, a dummy for large company size and nine dummies for industry. Male and female wage equations include also a dummy for parenthood and a dummy for marriage (as they have an opposite effect on wages of men and women).

The Duncan and Hoffman specification has the advantage of being able to test two competing theories (Hartog and Oosterbeek, 1988). If human capital theory holds, wages are not influenced by particular job requirements and therefore years of education should be rewarded equally, regardless of whether they are matched with, over- or under- the job requirements. In this case, $\beta_1 = \beta_2 = -\beta_3$ and equation (4) would be reduced to the standard Mincerian wage equation (2). The last but one row of Table 5, which states the results of equations (2) and (4) using an OLS regression, shows that based on the computed F-statistics this null hypothesis that $\beta_1 = \beta_2 = -\beta_3$ can be rejected regardless of which approach and dataset is used.

According to job competition theory (Thurow, 1975), wages are related to jobs rather than to workers, and thus only the years of education required by the job in question are rewarded. In this case $\beta_2 = \beta_3 = 0$ and only β_1 in equation (4) would be non-negative. The last row of Table 5 shows that based on the computed F-statistics this null hypothesis that $\beta_2 = \beta_3 = 0$ can be also rejected regardless of which approach and dataset is used.

The Mincerian-type wage equation shows a 6.4 percent wage increase for every additional year of actual education using the SILC data, and a 5.2 percent increase using the PIAAC data. The returns to actual years of education are higher for women than for men. The Duncan and Hoffman model confirms expectations based on existing empirical research (e.g. Alba-Ramírez, 1993; Nieto, 2014, for Spain) that returns to required years of education are higher than returns to actual years of education attained. Moreover, returns to years of overeducation are positive but smaller than returns to required years of education ($\beta_1 > \beta_2 > 0$) and returns to years of undereducation are negative but smaller than returns to required years of education ($\beta_1 > -\beta_3 > 0$).

The results with the statistical approach using the SILC and PIAAC data differ only in the wage premium for years of overeducation. The self-declared approach using PIAAC data results in 3.3 percentage point smaller returns to required years of education, a 1.2 percentage point higher wage premium from years of overeducation and a c. 2.6 percentage point smaller wage penalty from years of undereducation than under the statistical approach. The most obvious gender difference is that the wage penalty from years of undereducation is almost twice as high for women than for men.

Conclusion

Three approaches to measuring over- and undereducation – defined as a mismatch between the educational level required by an occupation and its workers' actual educational attainment – have been applied in the literature: the normative approach, based on an exogenous definition determined by job analysts, the statistical approach, based on workers' observed educational attainment levels, and the self-declared approach, based on workers' self-evaluation of the education level required in their occupation.

In this study I have applied all three approaches to Czech data and, like existing studies for other countries that also applied more than one approach, have come to the conclusion that the indices of over- and undereducation vary substantially according to the approach applied: the share of overeducation reported differs by up to 17 percentage points and the share of undereducation by up to 13 percentage points depending on the

approach applied. There is no ideal approach to educational mismatch analysis: all the approaches used lead to different results, and all have their advantages and disadvantages.

This study has also shown that different data sources lead to different conclusions. I applied data from the national Labour force Survey (LFS), the national Statistics on Income and Living Conditions (SILC), and the Programme for the International Assessment of Adult Competencies (PIAAC) in my measurements of educational mismatch. Of these, only the PIAAC data source can be used for the self-declared approach. The incidence of both over- and undereducation differs by up to 4 percentage points when the same approach is used, but a different data source. Of the three data sources used, the PIAAC data yield the highest shares of overeducation and the LFS data report the lowest shares of undereducation.

I further used the SILC and PIAAC data to analyse the determinants of educational mismatch and wage effects. The impacts of various characteristics on the likelihood of being over- and undereducated vary substantially across both data sources and the three measurement approaches, in some cases resulting in opposite effects. Even the effect of gender differs across the approaches: men are less likely to be overeducated than women under the normative approach, but more likely under the statistical approach. Moreover, the effects are more pronounced when using PIAAC data than SILC data.

The wage effects of educational mismatch are in accordance with previous findings for other countries. Overeducated workers have lower wages and undereducated workers have higher wages compared to workers with the same education in jobs that match their level of education. However, these estimates also vary in their size depending on the approach and data source used. PIAAC data report a higher wage penalty for being overeducated than SILC data, by up to 4 percentage points, and a higher wage premium for being undereducated, by about 2 percentage points. Both the wage penalty and the wage premium are shown to be highest under the normative approach.

Overeducated workers earn more than their co-workers with exactly the required education level, and undereducated workers earn less. Returns to required years of education are higher than returns to years of education attained. Moreover, returns to years of surplus education are positive but lower than the returns to required years of education and, vice versa, returns to years of deficit education are negative but this penalty is lower than the returns to required years of education. The effects are qualitatively the same regardless of which data source or approach is applied, although the PIAAC data indicates lower returns to years of surplus education than the SILC data.

The results of all the wage models, regardless of the data sources and approaches applied, indicate that both the human capital model, which assumes equal returns to required, surplus and deficit years of education, and the job competition model, which assumes no returns to surplus or deficit years of education, can be rejected. The findings suggest that wages are neither solely related to a worker's education level nor solely related to the level of education a job requires, but somewhere in the middle.

Appropriate next steps in the analysis of educational mismatch in the Czech labour market would be to include heterogeneity in skills among workers with the same educa-

tion level (and field of study), as existing empirical research in more analysed countries has done. This might help to address another potentially important feature of tertiary education in the CR in particular: the quality and diversity of universities.

Furthermore, it would be desirable not only to extend the analysis to examine changes in educational mismatch over time, but also to focus on the persistence of educational mismatch. Empirical studies have typically confirmed that years of work experience have a robust negative impact on the probability of being overeducated. Workers with lack of experience might accept overeducation in order to gain more experience and consequently to improve their job level, meaning that this negative effect should be temporary. This negative effect was seen only in few of the models tested in this study, suggesting that overeducation may be a persistent rather than temporary phenomenon, mainly among young workers.

Disclosure statement: No potential conflict of interest was reported by the author.

References

ALBA-RAMÍREZ, A. (1993). Mismatch in the Spanish labour market. Overeducation? *Journal of Human Resources*. 28, pp. 259–278.

BAUER, T. (2002). Educational mismatch and wages: a panel analysis. *Economics of Education Review*. 21, pp. 221–229.

BEVAN, S. & COWLING, M. (2007). Job matching in the UK and Europe. Research report 25, The Work Foundation. http://dera.ioe.ac.uk/7125/1/research-report-25-job-matching-in-the-uk-and-europe.pdf

CHEVALIER, A. (2003). Measuring over-education. *Economica*. 70, pp. 509–531.

DUNCAN, G. J. & HOFFMAN, S. D. (1981). The incidence and wage effects of over-education. *Economics of Education Review*. 1, pp. 75–86. DOI: 10.1016/0272-7757(81)90028-5

FREEMAN, R. (1976). *The overeducated American*. New York: Academic Press.

GOTTSCHALK, P. & HANSEN, M. (2003). Is the proportion of college workers in non-college jobs increasing? *Journal of Labor Economics*. 21, pp. 449–471. DOI: 10.1086/345564

HARTOG, J. & OOSTERBEEK, H. (1988). Education, allocation and earnings in the Netherlands: overschooling? *Economics of Education Review*. 7, pp. 185–194. DOI: 10.1016/0272-7757(88)90043-X

HERNÁNDEZ, L. & SERRANO, L. (2012). Overeducation and its effects on wages: a closer look at the Spanish regions. *Investigaciones Regionales*. 2012, pp. 59–90.

JEONG, B., KEJAK, M. & VINOGRADOV, V. (2008). Changing composition of human capital. The Czech Republic, Hungary and Poland. *Economics of Transition*. 16, pp. 247–271. DOI: 10.1111/j.1468-0351.2008.00320.x

KARAKAYA, G., PLASMAN, R. & RYCX, F. (2007). Overeducation on the Belgian labour market: evaluation and analysis of the explanatory factors through two types of approaches. *Compare: A Journal of Comparative and International Education*. 37, pp. 513–532.

KIKER, B. F., SANTOS, M. C., & DE OLIVEIRA, M. M. (1997). Overeducation and undereducation: evidence for Portugal. *Economics of Education Review*. 16, pp. 111–125. DOI: 10.1016/S0272-7757(96)00040-4

MCGOLDRICK, K. M. & ROBST, J. (1996). Gender differences in overeducation: a test of the theory of differential overqualification. *The American Economic Review*. 86, pp. 280–84.

MINCER, A. (1974). *Schooling, experience and earnings*. New York: NBER Press.

MYSÍKOVÁ, M. (2011). EU-SILC and its methodological pitfalls: international comparability and income variables. *Data a výzkum – CSA Info*. 5, pp. 147–170 (in Czech).

NIETO, S. (2014). Overeducation, skills and wage penalty: Evidence for Spain using PIAAC data. IREA Working Paper 2014/11. Barcelona: Research Institute of Applied Economics.

OECD (2007). *International migration outlook: SOPEMI 2007 edition*. Paris: OECD. DOI: 10.1787/migr_outlook-2007-sum-is

RAŠOVEC, T. & VAVŘINOVÁ, T. (2014). Skills and educational mismatch in the Czech Republic: comparison of different approaches applied on PIAAC. *Statistika* 94, pp. 58–79.

ROSEN, S. (1972). Learning and experience in the labor market. *Journal of Human Resources*. 7, pp. 326–342. DOI: 10.2307/145087

RUBB, S. (2003). Overeducation: a short or long run phenomenon for individuals? *Economics of Education Review*. 22, pp. 389–394. DOI: 10.1016/S0272-7757(02)00052-3

THUROW, L. C. (1975). *Generating inequality*. New York: Basic Books.

VAN DER MEER, P. & WIELERS, R. (1996). Educational credentials and trust in the labour market. *Kyklos*. 49, pp. 29–46. DOI: 10.1111/j.1467-6435.1996.tb01383.x

VEČERNÍK, J. (2013). The changing role of education in the distribution of earnings and household income. The Czech Republic, 1988–2009. *Economics of Transition*. 21, pp. 111–133. DOI: 10.1111/ecot.12002

VERDUGO, R. & VERDUGO, N. (1989). The impact of surplus schooling on earnings: some additional findings. *Journal of Human Resources*. 24, pp. 629–643. DOI: 10.2307/145998