Health Outcomes of Educational Mismatch: A Direct Effect or a Matter of Reporting Heterogeneity? Evidence from the Russian Federation

Mariia Vasiakina[1] and Silvana Robone[1234]

[1] Department of Economics, University of Insubria, Italy

[2] HEDG - Health, Econometrics and Data Group, University of York, UK

[3] Dondena Centre for Research on Social Dynamics and Public Policy, Bocconi University, Italy

[4] Hospinnomics, Paris School of Economics

Abstract

Our paper aims at contributing to the existing literature on educational mismatch by investigating the impact of over-education and under-education on some subjective and objective health outcomes of employees - self-assessed health (SAH) and hypertension, respectively - and conducting a longitudinal gender-specific analysis on data from the Russia Longitudinal Monitoring Survey (RLMS-HSE) (2000-2014). We estimate dynamic correlated random effects ordered probit and probit models for SAH and hypertension, respectively. The results provide evidence that both over-education and under-education are related to our objective health measure but the effect can only be observed for the male sub-sample. On the contrary, no impact of educational mismatch on SAH is observed for both gender groups. Due to lack of consistency between the estimates for subjective and objective health outcomes, we make the hypothesis that SAH is affected by the issue of reporting heterogeneity. To test this hypothesis, the RLMS-HSE (2005) is merged with externally collected anchoring vignettes for Russia from the World Health Survey (2003). The findings from estimated ordered probit and hierarchical ordered probit (HOPIT) models suggest that reporting heterogeneity might be the issue why we observe the effect of mismatch on the objective measure of health, but not on the self-reported one.

Keywords: educational mismatch, over-education, under-education, self-assessed health, EQ-5D, panel data analysis, RLMS-HSE, World Health Survey, self-reported bias, anchoring vignettes, HOPIT model

JEL Codes: I10, I26, I29, J01, J24, J80, J81

Corresponding author: Mariia Vasiakina, University of Insubria, Department of Economics, Via Monte Generoso 71, Varese, 21100. Email: mvasiakina@uninsubria.it

1. Introduction

Over the past decades the policy making institutions in low- and middle-income countries have persistently considered the better educated working population as a precondition for sustainable economic growth and development. However, recent studies have argued that economic growth is related to an appropriate distribution of employees over occupations with respect to their knowledge and skills, e.g. highly skilled positions are occupied by well-educated labor force, rather than to the higher educational attainment of the labor force (Handel et al., 2016).

Mapping of the outcomes of national educational systems with their labour market needs has indicated the presence of structural inefficiencies which take a form of educational mismatch. Being caused by either labour market imbalances, e.g. excess labour demand, risks and uncertainties, or institutional factors, e.g. technological change and short-term recessions, the issue of educational mismatch has been observed in both developed and developing economies and remains persistent over time (Davia et al., 2017; Ghaffarzadegan et al., 2017). For instance, evidence for the European countries suggests that the share of horizontally mismatched workers (i.e. whose field of study and occupation differ significantly) varies between 20 and 50 percent, while vertical mismatch (i.e. difference between the actual level of education which an individual has and the one which is required for his/her occupation) accounts for 15-35 percent in these countries (Morgado et al., 2016).

Recent research on educational mismatch has mainly focused on labour market outcomes. In particular, this issue has been studied in a context of wages and returns to education (Iriondo and Perez-Amaral, 2016; Pecoraro, 2016; Montt, 2017; Romero et al., 2017; Sellami et al., 2017) and related to firm productivity (Kampelmann and Rycx, 2012), job satisfaction (Badillo-Amador and Vila, 2013), and career dynamics (Mavromaras and McGuinness, 2012; Baert et al., 2013; Kiersztyn, 2013; Meroni and Vera-Toscano, 2017). However, there exist a few studies which investigate the association between educational mismatch and health-related outcomes of the labor force. They are mainly dedicated to the impact of over-education on depressive symptoms and provide evidence in favor of diminishing mental health returns to education (Bracke et al., 2014) and higher risk of mortality among over-educated respondents (Garcy, 2015).

This study aims at contributing to a limited number of papers on the health outcomes of educational mismatch and differs from the previous literature due to investigating the impact of both types of vertical educational mismatch - over-education and under-education - on both subjective and objective measures of health - self-assessed health and hypertension, respectively. First, we conduct a longitudinal gender-specific analysis and estimate dynamic correlated random effects ordered probit and probit models on a sample taken from the Russia Longitudinal Monitoring Survey (2000-2014). Thereafter, when we consider the subjective measure of health, we question the validity of our results by controlling for reporting heterogeneity. Our analysis is original since, to the best of our knowledge, this is the first study which considers the issue of reporting heterogeneity in a context of educational mismatch. To test for the reporting heterogeneity bias, we merge the RLMS-HSE (2005) with externally collected anchoring vignettes for Russia from the World Health Survey (2003) and compare the results for ordered probit and hierarchical ordered probit (HOPIT) models.

In addition, our analysis is related to a particular type of country – an economy in transition. Since transition economies have been mainly overlooked for the analysis on educational mismatch (Kyui, 2010; Kupets, 2016a), this study aims to fill the existing gap in the literature and pays attention to one of the countries where 'large imbalances between the supply and demand for skills… are driven by rapid economic restructuring, misalignment of the education system with labor market needs, and underdeveloped adult education and training systems' (Kupets; 2016b, p.1). In addition, Russia belongs to the group of countries with the highest prevalence of hypertension and, consequently, ischemic heart disease and cerebrovascular disease mortality rates (WHO, 2013) what justifies the choice of the objective health outcome for our study.

Our findings provide evidence that both types of vertical educational mismatch - over-education and under-education – are related to health of the Russian working population. However, the effect is gender-specific and clearly visible only for hypertension in the male sub-sample of employees. More precisely, under-educated men are less likely to be hypertensive than their matched counterparts, while the opposite can be observed for the over-educated group of Russian male employees. No effect of educational mismatch on both hypertension and self-assessed health is found for female employees in Russia. After adjusting for reporting heterogeneity, we obtain results which are partially consistent with our findings for the objective health measure, i.e. under-education remains positively associated with better health in the male sub-sample.

2. Research Methodology

2.1. Data, Sample and Variables

For the purpose of this study, we use the data from the 9th -23rd waves (2000-2014) of the Russia Longitudinal Monitoring Survey (RLMS-HSE) - a series of nationally representative surveys which cover the years of 1994-2017. Since it has been designed in line with the well-established longitudinal national surveys, e.g. the British Household Panel Survey (BHPS), the German Socio-Economic Panel (GSOEP) etc., the RLMS-HSE contains a wide range of data on health, education, work and welfare related topics which are collected at both individual and household levels.

We conduct our analysis on an unbalanced sample of currently working Russian employees aged 19 to 54. Hence, self-employed, entrepreneurs, retired, and unemployed respondents, as well as those being on leave at the moment of the survey are excluded from the analysis. We also exclude army forces due to the very specific tasks of this occupational group. Overall, our final sample includes 50,827 observations spread over 14 years.

Our dependent variables - the objective and subjective health measures - are represented by hypertension and self-assessed health (SAH), respectively. The variable of hypertension is defined in the RLMS-HSE by the question ‘Does your medical doctor say that you have hypertension?’ and measured on a binary scale (0 if respondent is diagnosed with hypertension and 1 otherwise). SAH is determined by the question ‘How would you evaluate your health?’ and measured on a 5-point categorical scale, where 1 equals to ‘very good’ and 5 equals to ‘very bad’ health. We rescale our SAH variable in order to be increasing in good health (from ‘very bad’ to ‘very good’) and reduce it to 3 categories due to the very small number of observations in the extreme categories. Hence, ‘very bad’ and ‘bad’ health outcomes are combined under a ‘very bad/bad’ category, while ‘very good’ and ‘good’ health responses – under a ‘good/very good’ one, respectively.

Our main explanatory variables - over-education and under-education – are dummy variables. They are constructed with respect to the empirical method of Verdugo and Verdugo (1989) which has been widely implemented in many studies (see McGuinness et al., 2017 for a systematic literature review on educational mismatch). According to this method, the mean value of years of completed education (computed by employees in each occupation) is compared with the actual number of years of completed education which an individual (employed in a particular occupation) has. If the actual number of years of education exceeds (is lower than) the mean value plus (minus) one standard deviation, this individual is classified as over-educated (under-educated). If the actual number of years of education belongs to this interval, the individual is treated as vertically matched in terms of education. In our analysis, matched respondents are taken as a reference category. We also control for a set of standard sociodemographic (e.g. age, marital status, educational attainment, income), work-related (e.g. working schedule, occupational class), and other (e.g. regional and year dummies) covariates (a full description of variables is provided in appendix, see Table A1).

Table 1 *- Descriptive statistics*

	Female sub-sample (n=27,788)				Male sub-sample (n=23,039)			
Variable	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max
Sociodemographic controls:								
Age	38.742	9.431	19	54	36.794	9.513	19	54
Single	0.402	0.490	0	1	0.270	0.444	0	1
12+ years of education	0.683	0.466	0	1	0.539	0.499	0	1
Income (ln)	9.039	0.934	5.298	11.451	9.378	0.892	5.298	11.462
Work-related controls:								
Part-time work	0.110	0.313	0	1	0.034	0.181	0	1
Over-time work	0.349	0.477	0	1	0.490	0.500	0	1
White collar	0.803	0.398	0	1	0.322	0.467	0	1
Main explanatory variables:								
Over-education	0.143	0.350	0	1	0.158	0.365	0	1
Under-education	0.147	0.355	0	1	0.133	0.339	0	1
Dependent variables:								
‘Very bad/bad‘ health	0.051	0.220	0	1	0.026	0.158	0	1
‘Fair’ health	0.618	0.486	0	1	0.497	0.5000	0	1
‘Good/very good' health	0.331	0.471	0	1	0.477	0.499	0	1
No hypertension	0.693	0.461	0	1	0.778	0.416	0	1

Table 1 contains descriptive statistics for the female and male sub-samples. The actual level of respondents’ education exceeds the one which is required for their occupation in 15.8 percent of cases for Russian men, while for women this share equals to 14 percent. About 14.7 and 13.3 percent of women and men in the sample are classified as under-educated, respectively. In addition, both gender groups tend to stick to the ‘fair’ category in self-assessment of their health (i.e. 62 vs 50 percent of responses among female and male sub-samples, respectively), while around 33 and 48 percent of women and men report ‘good/very good’ health.[1] Finally, around 31 percent of female and 22 percent of male employees are diagnosed with hypertension.

2.2. Estimation strategy

We estimate ordered probit and probit models for SAH and hypertension, respectively. To control for state dependence, we estimate dynamic specifications by including both one-year lag and initial value of

[1] The distribution of responses over SAH categories in Russia significantly differs from that of the EU-28. On average, 65 and 70 percent of women and men report ‘good/very good’ health in Europe, respectively, while only 25 and 22 percent of female and male respondents choose the ‘fair’ health category amongst others (Eurostat, 2016).

our dependent variables in the models (Wooldridge, 2005).[2] In addition, we use the Mundlak (1978) correction to account for correlated individual effects in the models which are represented by a function of within-individual means of the time-varying regressors. Our analysis is stratified by gender as done in the previous literature (Cottini, 2012; Campos-Serna et al., 2013).

Since SAH is a self-reported measure, instead of the latent level of health SAH_{it}^{*} we can only observe an indicator of the category SAH_{it} which our latent variable belongs to:

$$\mathrm{SAH_{it}} = \mathrm{j} \text{ if } \mu_{\mathrm{j}-1} < \mathrm{SAH_{it}^{*}} < \mu_{\mathrm{j}}\,,\ \mathrm{j} = 1 \ldots 2,\ \text{ where } \mu_0 = -\infty\,, \mu_{\mathrm{j}-1} \le \mu_{\mathrm{j}}\,, \mu_2 = +\infty \qquad (1)$$

Hence, the latent variable SAH_{it}^{*} can also be represented as a linear function of a vector of the main explanatory variables, socioeconomic, work-related and other covariates, and a random error term ε_{it} which is assumed to be normally distributed.

3. Results

Results of gender-specific estimation of dynamic ordered probit and probit models for SAH and hypertension, respectively, are displayed in Table 2. We observe two types of asymmetric effects in our estimates. First of all, the effects of the regressors on the same health outcome (i.e. either subjective or objective one) vary with respect to gender. Secondly, there are substantial differences in the effects of the same regressors on the subjective and objective health outcomes within the same gender group.

Our findings reveal no effect of any type of educational mismatch on SAH for both gender groups. This result is in line with the evidence from other studies (Hultin et al., 2016; Zhu and Chen, 2016). However, over-education tends to increase the risk of hypertension in the male sub-sample (the coefficient for over-education on reporting ‘no hypertension’ is negative and statistically significant). This is in line with the previous evidence regarding the negative impact of over-education on mental health and depressive symptoms (Mossakowski, 2011; Hardie, 2014; Milner et al., 2017). On the contrary, under-educated men in Russia are less likely to be hypertensive (the coefficient for under-education on reporting ‘no hypertension’ is positive and statistically significant). This result corresponds to the findings that under-education is related to challenging tasks and, as a result, improves cognitive resilience of employees (de Grip et al., 2007) and reduces risk of mortality (Garcy, 2015).

[2] Since we estimate dynamic models, the year of 2000 will be present in the analysis only in the form of initial value for dependent variables.

Table 2 – *Results of estimation: dynamic ordered probit and probit*

	SAH				No hypertension			
	Women		Men		Women		Men	
Variables	Coef.	St. Err.	Coef.	St. Err.	Coef.	St. Err.	Coef.	St. Err.
Sociodemographic controls:								
Age	-0.019*	0.010	-0.005	0.012	-0.012	0.014	-0.011	0.016
Age^2	-0.000	0.000	-0.000**	0.000	-0.001***	0.000	-0.000	0.000
Single	-0.031	0.024	0.007	0.031	-0.005	0.030	0.129***	0.041
12+ years of education	0.061**	0.030	-0.019	0.031	0.082**	0.038	0.043	0.041
Income (ln)	0.020	0.022	0.032	0.024	-0.025	0.027	-0.006	0.031
Work-related controls:								
Part-time work	-0.050	0.032	-0.035	0.061	-0.091**	0.044	-0.038	0.082
Over-time work	-0.019	0.023	-0.013	0.023	0.016	0.028	-0.027	0.030
White collar	0.096***	0.030	0.043	0.030	0.040	0.038	-0.077*	0.040
Main explanatory variables:								
Over-education	-0.004	0.032	0.008	0.035	0.009	0.042	-0.092**	0.044
Under-education	-0.024	0.034	-0.023	0.039	0.026	0.044	0.113**	0.053
State dependence:								
SAH in t-1:								
‘Fair’ health	0.616***	0.062	0.434***	0.081				
‘Good/very good' health	1.160***	0.073	0.878***	0.090				
SAH in t=1:								
‘Fair’ health	0.778***	0.068	0.748***	0.098				
‘Good/very good' health	1.536***	0.080	1.515***	0.106				
No hypertension in t-1					0.679***	0.036	0.648***	0.044
No hypertension in t=1					1.272***	0.052	1.388***	0.066
Log-Likelihood	-17211.521		-14513.965		-11181.578		-8477.325	

Coefficients were also estimated for regional dummies, year dummies, and regressors of within-individual means, but they are not reported here.
Robust standard errors are clustered at an individual level.
**p<0.10, **p<0.05, ***p<0.01*

Married men in Russia are more likely to suffer from hypertension, while earlier studies show that married individuals in Western Europe and the USA are better off in terms of both cardiovascular diseases and risk factors than their single counterparts (Manfredini et al., 2017; Schultz et al., 2017). Similar to the recent literature on the relationship between health and educational attainment (Fletcher, 2015; Lynch et al., 2016), our study reveals a statistically significant ‘protective’ effect of education on both SAH and hypertension for the female sub-sample.

Our findings also suggest that part-time work is positively related to hypertension for the Russian female employees (the coefficient for part-time on reporting ‘no hypertension’ is negative and statistically significant). This result is in line with recent studies which indicate that precarious employment is related to worse health outcomes of employees (Vancea and Utzet, 2017), in particular, putting a disproportionally large burden on women’s health (Menendez et al., 2007). Moreover, our analysis shows a positive and statistically significant relationship between the white collar position and better SAH in the female sub-sample. This is in line with previous literature which shows that blue collars in general (Ravesteijn et al., 2018) and female blue collars in particular (Esler et al., 2018) are worse off in terms of health than their white collar counterparts. White collar men, on the contrary, are more likely to be hypertensive. This might be explained by their physical strain due to excessive workload and low job control, which affect employees' health negatively (Ravesteijn et al., 2018).

Last but not least, our findings reveal a strong impact of state dependence on both subjective and objective health outcomes. We observe considerable estimates (both in terms of magnitude and statistical significance) for a one-year lag and initial value of the dependent variables over all specifications.

Table 3 displays the average marginal effects, representing the probability of reporting 'good/very good' health and 'no hypertension' status. While over-education increases the probability to be hypertensive by 1.5 percentage points for Russian men, under-educated male employees, on the contrary, experience a 'protective' effect against hypertension which equals to 1.9 percentage points.

***Table 3** – Average marginal effects*

	SAH				No hypertension			
	Women		Men		Women		Men	
Variables	dy/dx	St. Err.	dy/dx	St. Err.	dy/dx	St. Err.	dy/dx	St. Err.
Sociodemographic controls:								
Age	-0.005*	0.003	-0.001	0.004	-0.003	0.003	-0.002	0.003
Age^2	0.000	0.000	-0.000**	0.000	-0.000***	0.000	0.000	0.000
Single	-0.008	0.006	0.002	0.010	-0.001	0.006	0.021***	0.007
12+ years of education	0.016**	0.008	-0.006	0.010	0.017**	0.008	0.007	0.007
Income (ln)	0.005	0.006	0.010	0.007	-0.005	0.006	-0.001	0.005
Work-related controls:								
Part-time work	-0.013	0.009	-0.011	0.019	-0.018**	0.009	-0.006	0.014
Over-time work	-0.005	0.006	-0.004	0.007	0.003	0.006	-0.004	0.005
White collar	0.025***	0.008	0.013	0.009	0.008	0.008	-0.013*	0.007
Main explanatory variables:								
Over-education	-0.001	0.009	0.003	0.011	0.002	0.008	-0.015**	0.007
Under-education	-0.006	0.009	-0.007	0.012	0.005	0.009	0.019**	0.009

**p<0.10, **p<0.05, ***p<0.01*

Both the mixed evidence on the health outcomes of educational mismatch in the previous literature and the lack of consistency between the SAH and hypertension estimates in our analysis make us lean towards the view that SAH might be affected by the issue of reporting heterogeneity. In particular, the heterogeneity in the reporting styles of respondents might be the reason for observing a significant effect of educational mismatch on hypertension but not on SAH in the male sub-sample. In the following section we investigate this issue further.

4. Check for reporting heterogeneity: HOPIT estimation with externally collected vignettes

The issue of reporting heterogeneity - also known as differential item functioning (DIF) - is related to any self-assessment which is measured on a categorical scale, e.g. SAH, EQ-5D indicators, life satisfaction, health system responsiveness etc. (Rice et al., 2012; Bzostek et al., 2016; Knott et al., 2017; Rossouw et al., 2018). It arises from the fact that individuals vary in terms of understanding and using ordinal response categories and, as a result, placing the cut points between adjacent response categories differently (King and Wand, 2007; Grol-Prokopczyk et al., 2015).

The ordered choice models fail to address the issue of reporting heterogeneity and, consequently, may provide misleading results. An extension of the standard ordered probit model - the hierarchical ordered probit (HOPIT) model – can be adopted to adjust for this source of bias (Tandon et al., 2003). Along with a self-assessed outcome of interest, the HOPIT model requires to use anchoring vignettes. Anchoring vignettes are represented by a set of questions which are asked to respondents in addition to the self-assessment of a certain outcome and measured on the same categorical scale as the self-assessment they complement to, but they refer to a hypothetical individual and situation (van Soest et al., 2011; O'Doherty et al., 2017). The general specification of the HOPIT model can be written as follows:

$$H_{ik}^{v*} = \eta_k + \varepsilon_{ik.}^{v} \quad \varepsilon_{ik}^{v} \sim N\left(0, \sigma_v^2\right) \tag{2}$$

where H_{ik}^{v*} is a health outcome for vignette *k*, perceived by individual *i* and unobserved to the researcher*;* η_k is the mean of the underlying scale for the vignette *k*; ε_{ik}^{v} is an idiosyncratic error term.

$$h_{ik}^{v} = j, \; if \quad \mu_i^{j-1} \leqslant H_{ik}^{v*} \prec \mu_i^{j}, \tag{3}$$

with $\mu_i^0 = -\infty$ and $\mu_i^j = \infty$; where h_{ik}^{v} is the vignette rating on a j-point categorical scale observed instead of H_{ik}^{v*}.

$$\mu_i^1 = X_i\gamma^1 + u_i,$$

$$\mu_i^j = \mu_i^{j-1} + exp(X_i\gamma^j) \tag{4}$$

where $u_i \sim N\left(0, \sigma_u^2\right)$ is an unobserved individual specific random effect which is assumed to be independent of X_i and the other error terms in the model; γ^j and σ_u^2 are other parameters to be estimated. Hence, while the ordered probit model implies the cut points to be represented by fixed constants (which are common to all individuals), the HOPIT model treats them as functions of covariates X.

In order to 'work properly', vignettes should fulfil two assumptions, i.e. *response consistency* and *vignette equivalence*. *Response consistency* requires that an individual uses the response categories in the same way for both the self-assessment and the rating of the hypothetical situation described in the vignettes. *Vignette equivalence* states that 'respondents may differ with each other in how they perceive the level of the variable portrayed in each vignette, but any differences must be random and hence independent of the characteristic being measured' (King et al., 2004, p. 194).

To test our hypothesis about reporting heterogeneity bias in SAH we need relevant vignettes, but the RLMS-HSE does not contain them. Assuming that SAH is highly correlated with EQ-5D indicators (van Reenen and Oppe, 2015), we replace our original subjective health measure SAH with some items of the EQ-5D, which are pain, anxiety and depression (all self-reported) and focus our attention on the 14th wave of the RLMS-HSE (where EQ-5D questions were asked). Further, we adopt the procedure of the HOPIT estimation with externally collected vignettes (see Harris et al., 2015 for a detailed review of the procedure). The externally collected vignettes were taken from the Russian sample of the World Health Survey (2003).

When estimating the HOPIT models, our new dependent variables are specified as follows: 'Do you feel any pain?' and 'Do you feel any anxiety/depression?'. We rescale them from "bad" to "good health" outcomes in such way that the 1st and 2nd categories equal to 'acute' and 'some' levels of pain/anxiety/depression, while the 3rd one is related to the lack of disease. Since the vignettes in the World Health Survey (2003) are measured on a 5-point categorical scale (decreasing in terms of health), we provide correspondence between our self-assessments and vignettes assessments by rescaling the latter and reducing their number of categories from 5 to 3 (see appendix, Table A2). The distribution of vignette responses over sociodemographic characteristics clearly indicates the presence of reporting heterogeneity in our sample (see appendix, Figure A1-A3).

As a next step, we merge the 14th wave of the RLMS-HSE (2005) with the World Health Survey (2003) for Russia on the base of sociodemographic characteristics as age, gender, marital status, educational attainment, and occupation. Table A3 shows significant differences between the initial RLMS-HSE (2005) sample and the one which contains vignettes (obtained through merging) in terms of marital status, educational attainment, income, occupation and both types of educational mismatch. Therefore, we construct post-stratification weights and apply them to the sample which contains vignettes in order to make our datasets comparable (see Harris et al., 2015 for a detailed review of the procedure). As a result, the weighted sample with vignettes does not differ (in terms of means) at a 10-percent significance level from the RLMS-HSE (2005) with EQ-5D indicators (see appendix, Table A3 (columns with p-value)).

Since the HOPIT model arises from the ordered probit model, it can also be estimated by maximum likelihood techniques. However, the likelihood function will be slightly modified in this case, i.e. it will include two separate parts - one for self-assessments and another for vignette assessment, and post-stratification weights will be applied to the vignette assessment part of the likelihood function (Harris et al., 2015). Table 4 - 6 contain results for the ordered probit and the HOPIT models for pain, anxiety and depression.[3]

The estimates for over-education and under-education in the ordered probit model (Table 4) do not indicate any direct effect on pain for both gender groups. No effect of educational mismatch on pain is revealed after adjusting for reporting heterogeneity. However, in the first cut point equation we observe a statistically significant coefficient of -0.208 for under-education in the male sub-sample. This result suggests that under-educated men are more optimistic in their reports on pain than their matched counterparts, i.e. a negative sign of the coefficient means here that while choosing between 'extreme' and 'some' categories of pain under-educated men are more likely to choose the second option.

Table 5 shows that under-education affects both anxiety and the way respondents report it, however, these effects differ with respect to gender. More precisely, we observe a negative and statistically significant effect of under-education on anxiety for women in the ordered probit specification which becomes even bigger when we adjust for reporting heterogeneity (i.e. -0.170 vs -0.257 for ordered probit and HOPIT models, respectively). This disparity is explained by the fact that under-educated women, while choosing between 'extreme' and 'some' categories of anxiety, prefer the last one (the coefficient in the first cut point equation is negative and statistically significant) meaning that they are more optimistic than matched women in reporting their level of anxiety. In the male sub-sample, we do not observe any effect of under-education on anxiety in the ordered probit model. However, results from the HOPIT model suggest that under-educated men rate their level of anxiety more pessimistically than their matched counterparts (the coefficients in both cut point equations are positive and statistically significant). This justifies why, after adjusting for reporting heterogeneity, we find a statistically significant 'protective' effect of under-education on this EQ-5D measure (indicated by the coefficient of 0.285 in the main equation of the HOPIT model). The lack of consistency between our ordered probit and HOPIT estimates might be explained by the fact that the 'protective' effect of under-education on anxiety for Russian men is totally absorbed by their pessimistic way of reporting anxiety.

[3] The ordered probit estimates for anxiety and depression are equal in Table 5 and 6, since there are no separate variables for these EQ-5D indicators in the RLMS-HSE, i.e. they both are aggregated within the variable which is specified as follows: 'Do you feel any anxiety/depression?'. So we sequentially use this self-assessment with two sets of vignettes, i.e. one for anxiety and another for depression, in our HOPIT models.

Table 4 *– Pain: ordered probit and HOPIT*

	Women				Men			
	Ordered probit (n=1,644)		Weighted HOPIT (n=1,150)		Ordered probit (n=1,329)		Weighted HOPIT (n=983)	
Variables	Coef.	St. Err.	Coef.	St. Err.	Coef.	St. Err.	Coef.	St. Err.
Main equation								
Age	-0.087***	0.029	-0.211***	0.060	-0.010	0.033	-0.082	0.051
Age^2	0.001*	0.000	0.002***	0.001	-0.000	0.000	0.001	0.001
Single	-0.036	0.063	-0.162	0.132	0.060	0.100	0.116	0.159
12+ years of education	-0.076	0.076	-0.208	0.166	0.172*	0.090	-0.080	0.130
Income (ln)	0.086*	0.046	0.123	0.087	0.074	0.052	0.070	0.074
Part-time work	0.080	0.093	0.106	0.181	-0.305	0.208	-0.173	0.259
Over-time work	-0.027	0.066	-0.123	0.123	-0.028	0.069	-0.088	0.096
White collar	0.011	0.077	0.600***	0.182	0.083	0.084	0.086	0.124
Over-education	0.040	0.096	0.192	0.167	-0.051	0.100	-0.007	0.140
Under-education	-0.022	0.096	0.021	0.199	0.058	0.115	0.043	0.159
Constant			6.317	1.336			3.709	1.166
Cut point 1	-3.728	0.648			-2.217	0.775		
Cut point 2	-1.675	0.641			-0.215	0.771		
Cut point equation 1								
Age			-0.073**	0.030			-0.110***	0.032
Age^2			0.001***	0.000			0.002***	0.000
Single			-0.389***	0.069			0.111	0.096
12+ years of education			-0.470***	0.080			-0.684***	0.081
Income (ln)			0.047	0.044			0.072*	0.044
Part-time work			0.025	0.097			0.130	0.154
Over-time work			-0.001	0.064			-0.123**	0.059
White collar			0.414***	0.088			0.246***	0.076
Over-education			0.080	0.093			-0.024	0.089
Under-education			0.086	0.092			-0.208**	0.092
Constant			1.636	0.655			1.736	0.699
Cut point equation 2								
Age			-0.085**	0.042			-0.077**	0.036
Age^2			0.001**	0.001			0.001**	0.001
Single			0.000	0.093			0.038	0.107
12+ years of education			0.061	0.111			-0.205**	0.088
Income (ln)			-0.036	0.059			-0.001	0.049
Part-time work			0.121	0.131			-0.028	0.183
Over-time work			-0.043	0.086			-0.015	0.068
White collar			0.537***	0.115			-0.013	0.082
Over-education			-0.002	0.123			0.032	0.098
Under-education			0.038	0.134			-0.043	0.108
Constant			4.367	0.914			3.650	0.801
V1 constant			1.277***	0.057			0.930***	0.052
V2 constant			4.916***	0.099			3.179***	0.075
V3 constant			1.785***	0.061			1.226***	0.060

Coefficients were also estimated on the regional dummies (in the main equation) but they are not reported here.
**p<0.10, **p<0.05, ***p<0.01*

***Table 5** – Anxiety: ordered probit and HOPIT*

	Women				Men			
	Ordered probit (n=1,644)		Weighted HOPIT (n=1,150)		Ordered probit (n=1,329)		Weighted HOPIT (n=983)	
Variables	Coef.	St. Err.	Coef.	St. Err.	Coef.	St. Err.	Coef.	St. Err.
Main equation								
Age	-0.046*	0.028	0.089**	0.041	-0.036	0.035	-0.245***	0.046
Age^2	0.000	0.000	-0.002***	0.001	0.000	0.001	0.003***	0.001
Single	-0.135**	0.065	0.040	0.094	-0.139	0.101	-0.213	0.134
12+ years of education	-0.110	0.084	-0.078	0.121	0.214**	0.094	0.054	0.114
Income (ln)	0.067	0.048	0.002	0.063	0.041	0.057	-0.093	0.065
Part-time work	0.014	0.098	-0.138	0.129	-0.433**	0.209	-0.035	0.233
Over-time work	-0.060	0.068	-0.062	0.090	-0.221***	0.072	-0.182**	0.086
White collar	0.010	0.082	0.537***	0.133	0.035	0.088	0.167	0.110
Over-education	-0.077	0.096	-0.046	0.121	-0.033	0.109	0.026	0.122
Under-education	-0.170*	0.100	-0.257*	0.143	0.091	0.117	0.285**	0.141
Constant			0.331	0.921			8.332	1.041
Cut point 1	-3.216	0.648			-3.555	0.811		
Cut point 2	-0.879	0.646			-1.115	0.813		
Cut point equation 1								
Age			0.179***	0.031			0.323***	0.034
Age^2			-0.002***	0.000			-0.004***	0.000
Single			0.099	0.066			0.221**	0.094
12+ years of education			0.118	0.082			0.303	0.081
Income (ln)			0.038	0.042			-0.019	0.046
Part-time work			0.102	0.092			0.262	0.162
Over-time work			-0.022	0.063			-0.050	0.062
White collar			0.495***	0.088			-0.504***	0.077
Over-education			-0.010	0.087			-0.055	0.088
Under-education			-0.180*	0.099			0.188*	0.099
Constant			-4.404	0.678			-5.278	0.736
Cut point equation 2								
Age			0.149***	0.029			-0.233***	0.037
Age^2			-0.002***	0.000			0.003***	0.001
Single			0.246***	0.069			-0.168	0.103
12+ years of education			0.128	0.084			-0.041	0.089
Income (ln)			-0.077*	0.046			-0.110**	0.050
Part-time work			-0.157	0.100			0.264	0.188
Over-time work			-0.026	0.066			-0.033	0.068
White collar			0.310***	0.089			0.078	0.086
Over-education			-0.032	0.096			0.044	0.098
Under-education			-0.112	0.099			0.232**	0.109
Constant			-0.415	0.655			7.746	0.824
V1 constant			0.819***	0.038			0.641***	0.060
V2 constant			3.708***	0.093			3.242***	0.082
V3 constant			0.964***	0.052			1.105***	0.060

Coefficients were also estimated on the regional dummies (in the main equation) but they are not reported here.
$^*p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$

Table 6 – Depression: ordered probit and HOPIT

	Women				Men			
	Ordered probit (n=1,644)		Weighted HOPIT (n=1,150)		Ordered probit (n=1,329)		Weighted HOPIT (n=983)	
Variables	Coef.	St. Err.	Coef.	St. Err.	Coef.	St. Err.	Coef.	St. Err.
Main equation								
Age	-0.046*	0.028	0.284***	0.042	-0.036	0.035	-0.276***	0.054
Age^2	0.000	0.000	-0.004***	0.001	0.000	0.000	0.003***	0.001
Single	-0.135**	0.065	-0.234**	0.094	-0.139	0.101	-0.201	0.158
12+ years of education	-0.110	0.084	-0.097	0.121	0.214**	0.094	-0.004	0.134
Income (ln)	0.067	0.048	-0.083	0.064	0.041	0.057	-0.085	0.076
Part-time work	0.014	0.098	-0.121	0.130	-0.433**	0.209	-0.098	0.280
Over-time work	-0.060	0.068	-0.050	0.090	-0.221***	0.072	-0.225**	0.100
White collar	0.010	0.082	0.684***	0.132	0.035	0.088	-0.011	0.127
Over-education	-0.077	0.096	-0.029	0.121	-0.033	0.109	0.002	0.142
Under-education	-0.170*	0.100	-0.138	0.145	0.091	0.117	0.371**	0.165
Constant			-2.680	0.938			9.184	1.216
Cut point 1	-3.216	0.648			-3.555	0.811		
Cut point 2	-0.879	0.646			-1.115	0.813		
Cut point equation 1								
Age			0.119***	0.029			0.137***	0.035
Age^2			-0.002***	0.000			-0.002***	0.001
Single			-0.209***	0.066			0.341***	0.102
12+ years of education			-0.204***	0.078			0.260***	0.087
Income (ln)			0.010	0.041			0.028	0.049
Part-time work			0.067	0.091			0.104	0.171
Over-time work			-0.047	0.061			-0.074	0.066
White collar			0.315***	0.082			-0.393***	0.080
Over-education			-0.095	0.087			0.011	0.092
Under-education			-0.250***	0.095			0.119	0.108
Constant			-2.497	0.631			-2.648	0.779
Cut point equation 2								
Age			0.382***	0.031			-0.255***	0.040
Age^2			-0.005***	0.000			0.003***	0.001
Single			-0.010	0.065			-0.142	0.112
12+ years of education			0.164**	0.080			-0.127	0.095
Income (ln)			-0.170***	0.044			-0.107	0.053
Part-time work			-0.122	0.094			0.308	0.209
Over-time work			-0.006	0.063			-0.031	0.071
White collar			0.530***	0.084			-0.137	0.090
Over-education			0.002	0.091			0.021	0.105
Under-education			0.047	0.094			0.299***	0.114
Constant			-4.202	0.657			8.324	0.884
V1 constant			0.859***	0.039			0.835***	0.074
V2 constant			3.253***	0.079			3.574***	0.089
V3 constant			0.861***	0.051			1.319***	0.064

Coefficients were also estimated on the regional dummies (in the main equation) but they are not reported here.
**p<0.10, **p<0.05, ***p<0.01*

Finally, the ordered probit model in Table 6 shows that under-educated women are worse off in terms of depression. However, this direct effect disappears in the HOPIT specification due to the evidence that under-educated women, on average, are more positive in their assessments of depression than their matched counterparts (the coefficient in the first cut point equation is statistically significant and equals to -0.250). Similar to the results for anxiety, we do not find any statistically significant effect of under-education on depression for men in the ordered probit model. However, after adjusting for reporting heterogeneity, such effect becomes significant, indicating a 'protective' effect of under-education against depression in the male sub-sample (the coefficient is statistically significant and equals to 0.371 in the main equation). The 'protective' effect of under-education on depression is likely not to be observed in the ordered probit model because it tends to be nullified by a strong negative reporting style of Russian men about depression (the coefficient of 0.299 in the second cut point equation is statistically significant).

5. Discussion

This paper aims at contributing to the literature on socioeconomic determinants of health and estimates the impact of vertical educational mismatch - the divergence of the level of education which an individual has from that one which is required for the position he/she holds in the labour market – on both subjective and objective health outcomes of the Russian working population. In addition, in order to check for the presence of reporting heterogeneity bias in our self-assessed heath outcome, we estimate a HOPIT model with externally collected vignettes, as proposed by Harris et al. (2015). To the best of our knowledge, this is the first study which addresses the heterogeneity in reporting styles of individuals related to educational mismatch.

Although our findings vary with respect to gender (statistically significant effects are observed only in the male sub-sample), this analysis sheds light on some mechanisms which stand behind the impact of educational mismatch on the health of employees. In particular, when considering self-reported measures of health, after purging the data from reporting heterogeneity our findings suggest that under-educated men in Russia are less likely to suffer from anxiety and depression than their matched counterparts. However, these positive direct effects seem to be nullified by the pessimistic reporting styles of Russian men with regard to anxiety and depression. This might justify why we do not observe a statistically significant effect of under-education on anxiety and depression in the ordered probit models. Hence, reporting heterogeneity might also be a possible explanation for the lack of consistency between the estimates for SAH (where we do not find any effect of under-education) and non-hypertensive status (where the effect is positive and statistically significant) in the male sub-sample. Furthermore, if we assume that general health is made of two components, the physical and psychological one, the HOPIT model results suggest that under-education is relevant for the psychological component of health (proxied by anxiety and depression) rather than the physical one (proxied by pain). This is also in line with our findings on a 'protective' and statistically significant effect of under-education on hypertension which is, to some extent, a stress-related disease (Scalco et al., 2005).

Our study has some limitations. First of all, our models may suffer from omitted variable bias, since, due to data availability, we only take into account the vertical type of educational mismatch and completely ignore the horizontal one (divergence between the field of study and occupation of an individual). Secondly, since there are no precise vignettes for SAH in the RLMS-HSE, we have to replace our dependent variable SAH by some self-reported EQ-5D indicators and switch to a cross-sectional analysis in order to test for reporting heterogeneity. These technicalities affect the precision of our results.

In the future we could replicate our analysis by using precise vignettes for SAH, when they are made available. It could also be relevant to perform cross-country comparisons and extend the analysis to other European countries (both Western and Eastern European) in addition to the Russian case. Finally, educational mismatch could be considered in a context of adverse health behaviors in order to test whether over-education and under-education might be a trigger for smoking and drinking behavior patterns among employees.

From the policy point of view our study provides evidence in favor of those reforms which are addressed to reducing the number of over-educated employees in the labor market, since over-education is related to both the inefficient utilization of human capital of the Russian employees and the negative spillovers on their health status. This aim could partially be achieved if individuals and HR managers in the companies succeed in making mindful decisions while posting a position and hiring personnel, respectively. In addition, the policy makers' effort could be focused on making 'blue collar' occupations more attractive for employment by improving working conditions and providing competitive rewards. In the long term, this may nudge the high school graduates to choose employment in the blue collar occupations instead of applying for the higher professional education. Finally, the output of the higher professional institutions should be aligned with the needs of the labor market in Russia. Overall, these initiatives might help to overcome the negative impact of over-education on health and labor market outcomes of the Russian employees.

References

Badillo-Amador L., Vila L.E. (2013). Education and skill mismatches: wage and job satisfaction consequences. International Journal of Manpower, 34(5), pp. 416-428.

Baert S., Cockx B., Verhaest D. (2013). Overeducation at the start of the career: stepping stone or trap? *Labour Economics, 25*, pp. 123-140.Bracke P., Pattyn E., von dem Knesebeck O. (2013). Overeducation and depressive symptoms: diminishing mental health returns to education. *Sociology of Health & Illnes,35(8),* pp. 1242-1259.

Bracke P., van de Straat V., Missinne S. (2014). Education, mental health, and education-labor market misfit. *Journal of Health and Social Behavior, 55(4),* pp. 442-459.

Bzostek S., Sastry N., Goldman N., Pebley A., Duffy D. (2016). Using vignettes to rethink Latino-white disparities in self-rated health. *Social Science and Medicine, 149*, pp. 46-65.

Campos-Serna J., Ronda-Perez E., Artazcoz L., Moen B.E., Benavides F.G. (2013). Gender inequalities in occupational health related to the unequal distribution of working and employment conditions: a systematic review. *International Journal for Equity in Health 2013, 12,* pp. 1-18.

Cottini E. (2012), Is your job bad for your health? Explaining differences in health at work across gender. *International Journal of Manpower, 33(3),* pp. 301-321.

Davia M.A., McGuinness S., O'Connell P.J. (2017). Determinants of regional differences in rates of overeducation in *Europe. Social Science Research, 63*, pp. 67-80.

De Grip A., Bosma H., Willems D., van Boxtel M. (2007). Job-worker mismatch and cognitive decline. *IZA Discussion paper №2956.*

Elser H., Falconi A.M., Bass M., Cullen M.R. (2018). Blue collar work and women's health: a systematic review of the evidence from 1990 to 2015. *Social Science and Medicine - Population Health, 6*, pp. 195-244.

Eurostat (2016). Distribution_of_persons_aged_16_and_over_by_self-perceived_health_status,_by_sex,_2016 (%) (online data code: hlth_silc_10).Fiorentini G., Ragazzi G., Robone S. (2015). Are bad health and pain making us grumpy? An empirical evaluation of reporting heterogeneity in rating health system responsiveness. *Social Science and Medicine, 144*, pp. 48-58.

Fletcher J.M. (2015). New evidence of the effects of education on health in the US: compulsory schooling laws revisited. *Social Science and Medicine, 127*. pp. 101-107.

Garcy A.M. (2015). Educational mismatch and mortality among native-born workers in Sweden. A 19-year longitudinal study of 2.5 million over-educated, matched and under-educated individuals, 1990-2008. *Sociology of Health & Illness, 37(8)*, pp. 1314-1336.

Ghaffarzadegan N., Xue Y., Larson R.C. (2017). Work-education mismatch: an endogeneity theory of professionalization. *European Journal of Operational Research, 261*, pp. 1085-1097.

Grol-Prokopczyk H., Verdes-Tennant E., McEniry M., Ispany M. (2015). Promises and pitfalls of anchoring vignettes in health survey research. *Demography*, 52(5), pp. 1703-1728.

Handel M.J., Valerio A., Sanchez Puerta M.L. (2016). Accounting for Mismatch in Low- and Middle-Income Countries: Measurement, Magnitudes, and Explanations. Directions in Development. Washington, DC: World Bank.

Hardie J.H. (2014). The consequences of unrealized occupational goals in the transition to adulthood. *Social Science and Medicine, 48*, pp. 196-211.

Harris M.N., Knott R., Lorgelly P., Rice N. (2015). Survey self-assessments, reporting behavior and the use of externally collected vignettes. Bankwest Curtin Economics Centre Working Paper 15/8, Perth: Curtin University.

Hultin H., Lundberg M., Lundin A., Magnusson C. (2016). Do overeducated individuals have increased risks of ill health?: a Swedish population-based cohort study. *Sociology of Health & Illness, 38(6),* pp. 980-995.

Iriondo I., Perez-Amaral T. (2016). The effect of educational mismatch on wages in Europe. *Journal of Policy Modeling, 38*, pp. 304-323.

Kampelmann S., Rycx F. (2012). The impact of educational mismatch on firm productivity: evidence from linked panel data. *Economics of Education Review, 31*, pp. 918-931.

Kiersztyn A. (2013). Stuck in a mismatch? The persistence of overeducation during twenty years of the post-communist transition in Poland. *Economics of Education Review, 32*, pp. 78-91.

King, G., Murray C.J.L., Salomon S.A., Tandon A. (2004). Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review, 98(1)*, pp. 191-207.

King G., Wand J. (2007). Comparing incomparable survey responses: evaluating and selecting anchoring vignettes. *Political Analysis, 15*, pp. 46-66.

Knott R.J., Longelly P.K., Black N., Hollingsworth B. (2017). Differential item functioning in quality of life measurement: an analysis using anchoring vignettes. *Social Science and Medicine, 190*, pp. 247-255.

Kupets O. (2016a). Education-job mismatch in Ukraine: too many people with tertiary education or too many jobs for low-skilled? *Journal of Comparative Economics, 44*, pp. 125-147.

Kupets O. (2016b). Skill mismatch and overeducation in transition economies. *IZA World of Labour, 224*, pp. 1-10.

Kyui N. (2010). Returns to education and education-occupation mismatch within a transition economy. Empirical analysis for the Russian Federation. *Documents de travail du Centre d'Economie de la Sorbonne*, 2010.

Lynch J.L., von Hippel P.T. (2016). An education gradient in health, a health gradient in education, or a confounded gradient in both? *Social Science and Medicine, 154*. pp. 18-27.

Manfredini R., De Giorgi A., Tiseo R., Boari B., Cappadona R., Salmi R., Gallerani M., Signani F., Manfredini F., Mikhailidis D.P., Fabbian F. (2017). *Journal of Women's Health, 26(6)*. doi.org/10.1089/jwh.2016.6103

Mavromaras K., McGuinness S. (2012). Overskilling dynamics and educational pathways. *Economics of Education Review, 31*, pp. 619-628.

McGuinness S., Pouliakas K., Redmond P. (2017). Skills mismatch: concepts, measurement and policy approaches. *Journal of Economic Surveys*, pp. 1-31.

Menendez M., Benach J., Muntaner C., Amable M., O'Campo P. (2007). Is precarious employment more damaging to women's health than men's? *Social Science and Medicine, 64*, pp. 776-781.

Meroni E.C., Vera-Toscano E. (2017). The persistence of overeducation among recent graduates. *Labour Economics, 48*, pp. 120-143.

Milner A., Aitken Z., Kavanagh A., LaMontagne A.D., Petrie D. (2017). Status inconsistency and mental health: a random effects and instrumental variables analysis using 14 annual waves of cohort data. *Social Science and Medicine, 189*, pp. 129-137.

Montt G. (2017). Field-of-study mismatch and overqualification: labour market correlates and their wage penalty. *Journal of Labor Economics, 6(2)*, pp. 1-20.

Morgado A., Sequeira T.N., Santos M., Ferreira-Lopes A., Reis A.B. (2016). Measuring labour mismatch in Europe. *Social Indicators Research, 129*, pp. 161-179.

Mossakowski K.N. (2011). Unfulfilled expectations and symptoms of depression among young adults. *Social Science and Medicine, 73*, pp. 729-736.

Mundlak Y. (1978). On the pooling of time series and cross section data, *Econometrica, 46(1)*, pp. 69-85.

O'Doherty M.G., French D., Steptoe A., Kee F. (2017). Social capital, deprivation and self-rated health: Does reporting heterogeneity play a role? Results from the English Longitudinal Study of Ageing. *Social Science and Medicine, 179*, pp. 191-200.

Pecoraro M. (2016). The incidence and wage effects of overeducation using the vertical and horizontal mismatch in skills. *International Journal of Manpower, 37(3)*, pp. 536-555.

Ravesteijn B., van Kippersluis H., van Doorslaer E. (2017). The wear and tear on health: what is the role of occupation? *Health Economics*. doi.org/10.1002/hec.3563.

Rice N., Robone S., Smith P.C. (2012). Vignettes and health systems responsiveness in cross-country comparative analyses. *Journal of the Royal Statistical Society, 175(2)*, pp. 337-369.

Romero L.M., Huertas I.P.M., Jimenez M.M.S. (2017). Wage effects of cognitive skills and educational mismatch in Europe. *Journal of Policy Modeling, 39*, pp. 909-927.

Rossouw L., Bago d'Uva T., van Dooerslaer E. (2018). Poor health reporting? Using anchoring vignettes to uncover health disparities by wealth and race. *Demography, 55,* pp. 1935-1956.

Scalco A., Scalco M., Azul J., Lotufo Neto F. (2005). Hypertension and depression. *Clinics, 60(3)*, pp. 241-250.

Schultz W.M., Hayek S.S., Tahhan A.S., Ko Y., Sandesara P., Awad M., Mohammed K.H., Patel K., Yuan M., Zheng S., Topel M.L., Hartsfield J., Bhimani R., Varghese T., Kim J.H., Shaw L., Wilson P., Vaccarino V., Quyyumi A.A. (2017). Marital status and outcomes in patients with cardiovascular disease. *Journal of American Heart Association, 6*. doi: 10.1161/JAHA.117.005890

Sellami S., Verhaest D., Nonneman W., Van Trier W. (2017). The impact of educational mismatches on wages: the influence of measurement error and unobserved heterogeneity. *The B.E. Journal of Economic Analysis and Policy*.

Tandon, A., Murray, C. J. L., Salomon, J. A. and King, G. (2003). Statistical models for enhancing cross-population comparability. In Health Systems Performance Assessment: Debates, Methods and Empiricism (eds. C. J. L. Murray and D. B. Evans), pp. 727-746. Geneva: World Health Organization

Van Reenen M., Oppe M. (2015). EQ-5D-3L User Guide. Basic information on how to use EQ-5d-3L instrument. EuroQol Research Foundation.

Van Soest A., Delaney L., Harmon C., Kapteyn A., Smith J.P. (2011). Validating the use of anchoring vignettes for the correction of response scale differences in subjective questions. *Journal of the Royal Statistical Society, 174(3)*, pp. 575-595.

Vancea M., Utzet M. (2017). How unemployment and precarious employment affect the health of young people: A scoping study on social determinants. *Scandinavian Journal of Public Health, 45*, pp. 73-84.

Verdugo R., Verdugo N.T. (1989). The impact of surplus schooling on earnings: some additional findings, *Journal of Human Resources*, *24(4)*, pp. 629-643.

WHO Report (2013). A global brief on hypertension: silent killer, global public health crisis. World Health Organization, Geneva, Switzerland. Document number WHO/DCO/WHD/2013.2.

Wooldridge J.M. (2005). Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity, *Journal of Applied Econometrics, 20(1)*, pp. 39-54.

Zhu R., Chen L. (2016). Overeducation, overskilling and mental well-being. *BE Journal of Economic Analysis and Policy.*

Appendix

Table A1 – *Description of variables*

Variable	Definition
Dependent variables:	
Self-assessed health (SAH)	Categorical variable which is assessed on a 3-ladder scale. It equals to 1 if respondent reports 'Bad-' health status; it equals to 2 if respondent's health status is reported as 'Average'; and it equals to 3 if respondent assesses his health as 'Good+'.
Hypertension	Dummy variable which equals to 0 if respondent *is diagnosed* with hypertension and 1 otherwise.
Sociodemographic controls:	
Age	Respondent's age in years (calculated)
Age^2	Squared age
Single	Dummy variable where the 'single' category combines never married, divorced/separated, and widowed respondents, while married/cohabiting respondents are taken as a reference category.
12+ years of education	Dummy variable which is constructed with respect to the median-based measure of education where those respondents who have less than 12 years of completed education are taken as a reference category.
Average monthly income (ln)	Log of an average individual monthly income
Work-related controls:	
Part-time work	Dummy variable where the 'part-time work' category includes those respondents who work on average *less* than 35-40 hours/week, while employees who work exactly 35-40 hours/week are taken as a reference category.
Over-time work	Dummy variable where the 'over-time work' category includes those respondents who work on average *more* than 35-40 hours/week, while employees who work exactly 35-40 hours/week are taken as a reference category.
White collar	Dummy variable where the 'white collar' category combines managers, professionals, technicians and associate professionals, clerical support workers, and service and sales workers, while the 'blue collar' category - skilled agricultural, forestry and fishery workers, craft and related trades workers, plant and machine operators, and assemblers, and elementary occupations - is taken as a reference one.
Main explanatory variables:	
Over-education	Dummy variable where respondents are classified as over-educated if their number of years of completed education is *higher* than the mean number of years of education observed for their occupation, while matched respondents are taken as a reference category.
Under-education	Dummy variable where respondents are classified as under-educated if their number of years of completed education is *lower* than the mean number of years of education observed for their occupation, while matched respondents are taken as a reference category.
Regional dummies	A set of regional dummies which include Metropolitan area (i.e. Moscow and Saint Petersburg), North and Northern West, Center and Central Black Earth region, Volga-Viatskiy region, Northern Caucasus, Ural, Western Siberia, Eastern Siberia and Far East. Metropolitan area is taken as a reference category.
Year dummies	A set of year dummies which cover the period of 2001-2014. The year of 2001 is taken as a reference category.

***Table A2** – Description of vignettes*

Pain:

Vignette 1: [Mark] has pain in his knees, elbows, wrists and fingers, and the pain is present almost all the time. It gets worse during the first half of the day. Although medication helps, he feels uncomfortable when moving around, holding and lifting things. Overall in the last 30 days, how much of bodily aches or pains did [name of person] have?

1 - extreme | 2 - some | 3 - no pain

Vignette 2: [Katie] cannot remember when she last felt pain as this has not happened for the last several years now. She does not experience any pain even after hard physical labor or exercise. Overall in the last 30 days, how much of bodily aches or pains did [name of person] have?

1 - extreme | 2 - some | 3 - no pain

Vignette 3: [Laura] has a headache once a month that is relieved one hour after taking a pill. During the headache she can carry on with her day to day affairs. Overall in the last 30 days, how much of bodily aches or pains did [name of person] have?

1 - extreme | 2 - some | 3 - no pain

Anxiety:

Vignette 1: [Jan] feels nervous and anxious. He worries and thinks negatively about the future, but feels better in the company of people or when doing something that really interests him. When he is alone he tends to feel useless and empty. In the last 30 days, how much of a problem did [name of person] have with worry or anxiety?

1 - extreme | 2 - some | 3 - no anxiety

Vignette 2: [Ken] loves life and is happy all the time. He never worries or gets upset about anything and deals with things as they come. In the last 30 days, how much of a problem did [name of person] have with worry or anxiety?

1 - extreme | 2 - some | 3 - no anxiety

Vignette 3: [Henriette] enjoys her work and social activities and is generally satisfied with her life. She gets depressed every 3 weeks for a day or two and loses interest in what she usually enjoys but is able to carry on with her day to day activities. Overall in the last 30 days, how much of a problem did [name of person] have with distress, sadness or worry?

1 - extreme | 2 - some | 3 - no anxiety

Depression:

Vignette 1: [Jan] feels nervous and anxious. He worries and thinks negatively about the future, but feels better in the company of people or when doing something that really interests him. When he is alone he tends to feel useless and empty. Overall in the last 30 days, how much of a problem did [name of person] have with feeling sad, low, or depressed?

1 - extreme | 2 - some | 3 - no depression

Vignette 2: [Ken] loves life and is happy all the time. He never worries or gets upset about anything and deals with things as they come. Overall in the last 30 days, how much of a problem did [name of person] have with feeling sad, low, or depressed?

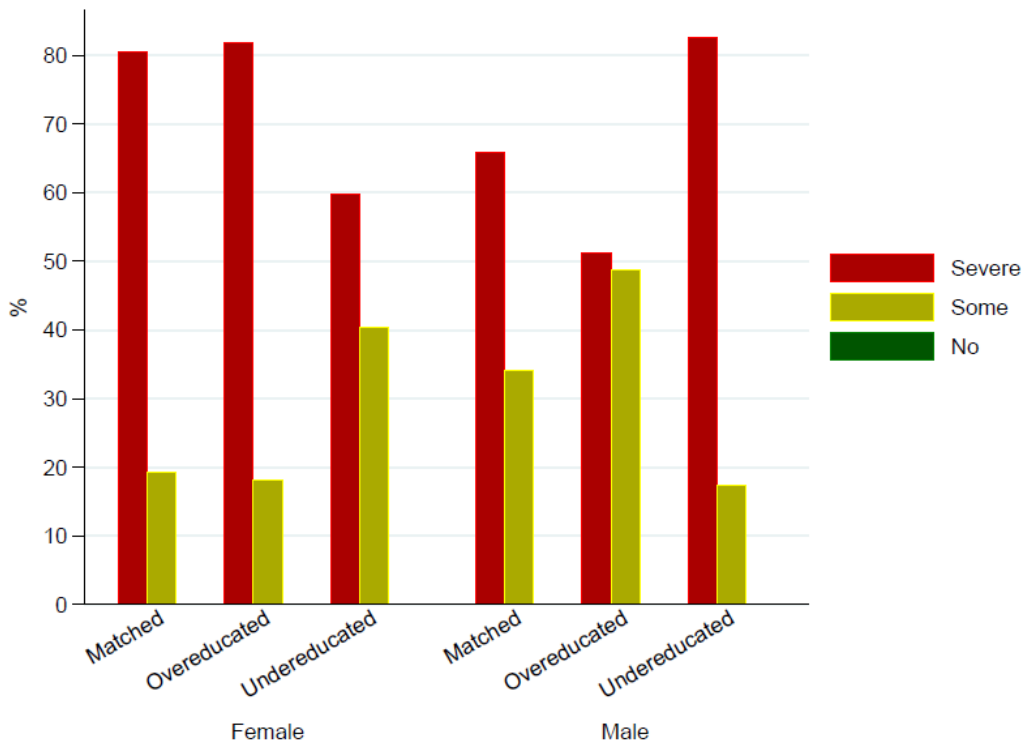
1 - extreme | 2 - some | 3 - no depression

Vignette 3: [Henriette] enjoys her work and social activities and is generally satisfied with her life. She gets depressed every 3 weeks for a day or two and loses interest in what she usually enjoys but is able to carry on with her day to day activities. Overall in the last 30 days, how much of a problem did [name of person] have with feeling sad, low, or depressed?

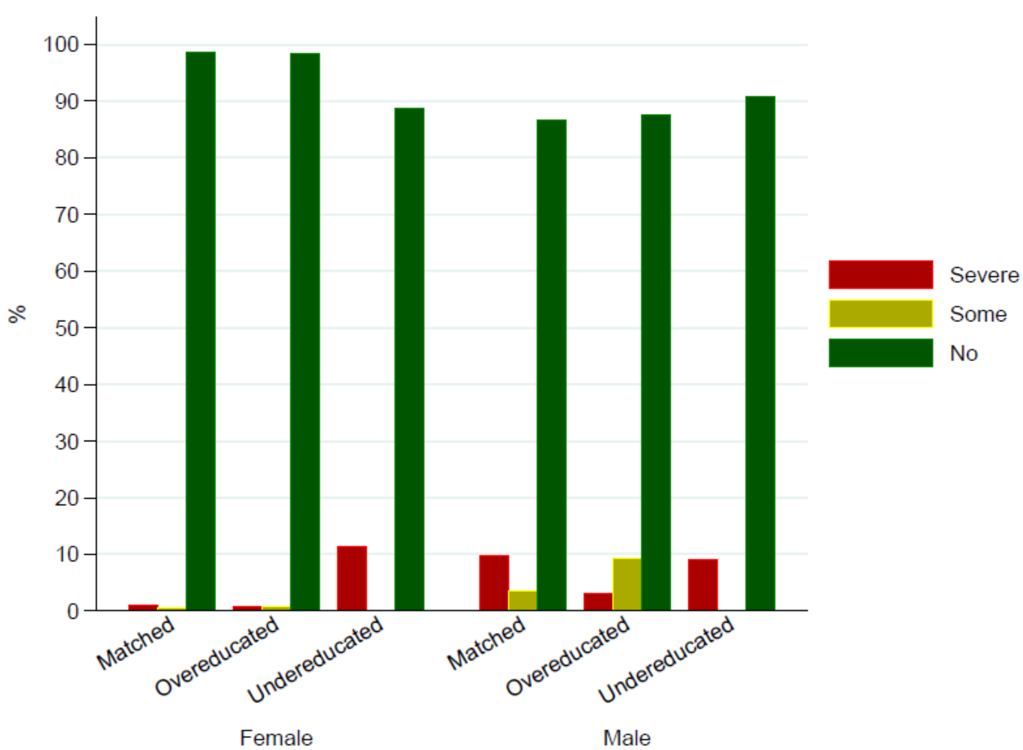
1 - extreme | 2 - some | 3 - no depression

__Figure A1__ - Distribution of vignettes responses over educational mismatch and gender: Pain

Vignette 1:



Vignette 2:



Vignette 3:

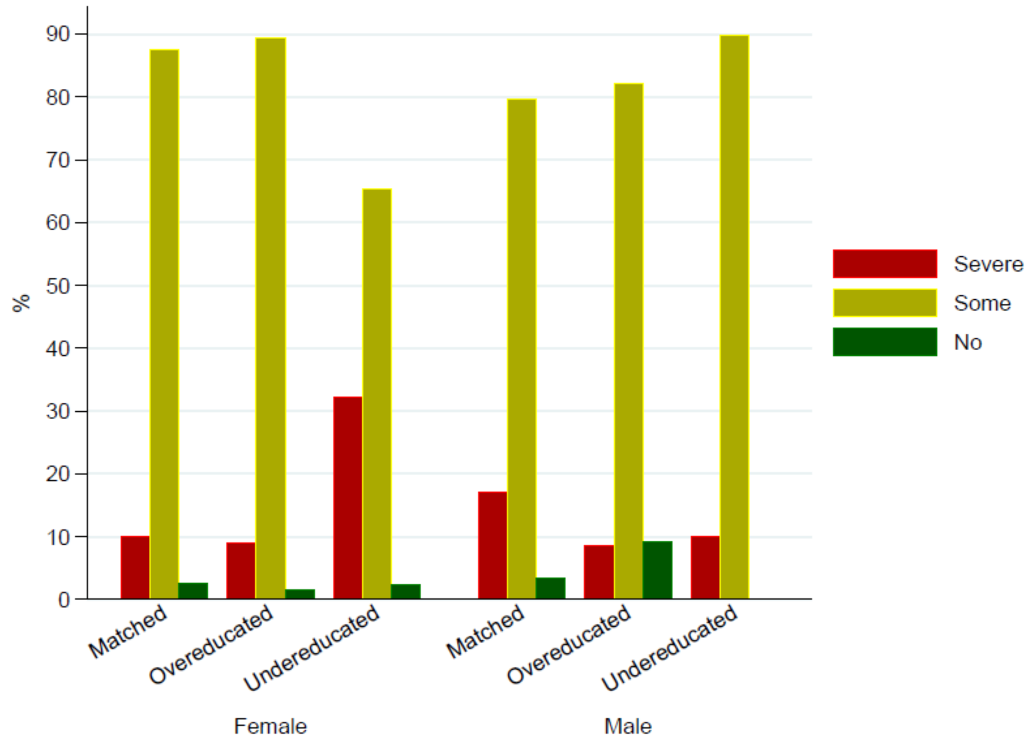
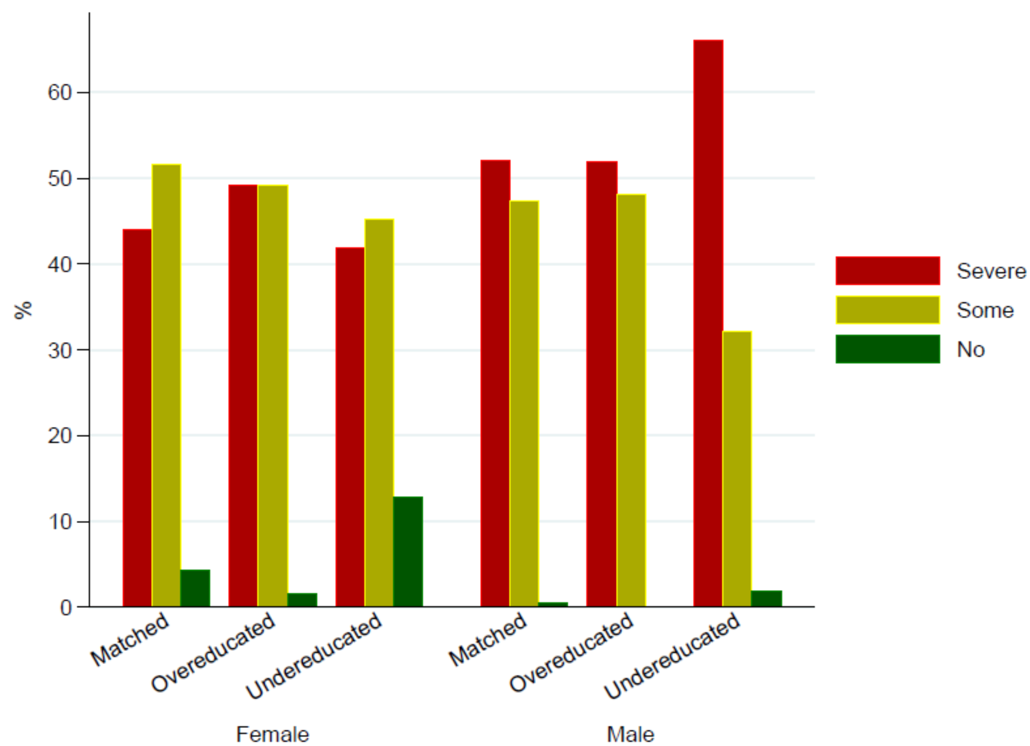
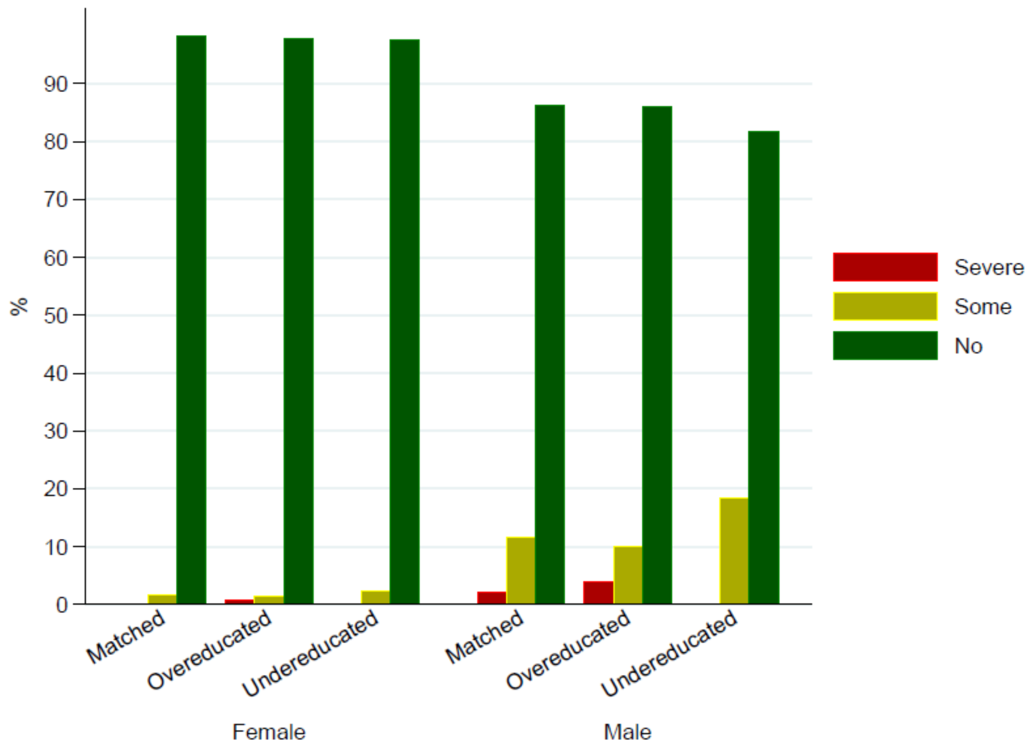


Figure A2 *- Distribution of vignettes responses over educational mismatch and gender: Anxiety*

Vignette 1:



Vignette 2:



Vignette 3:

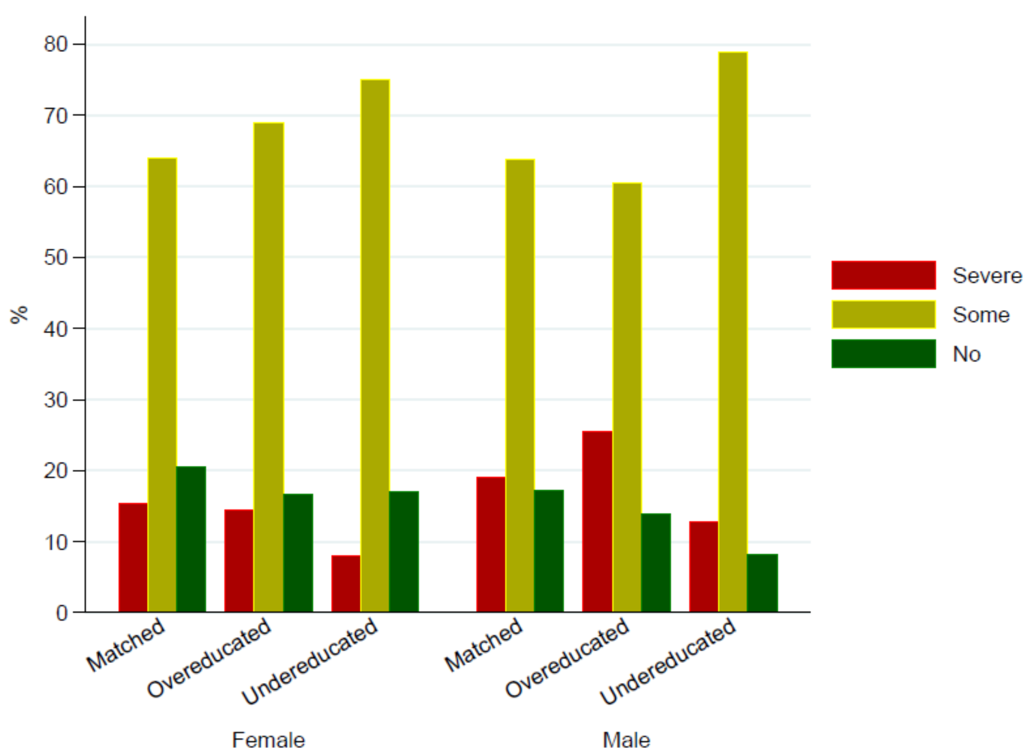
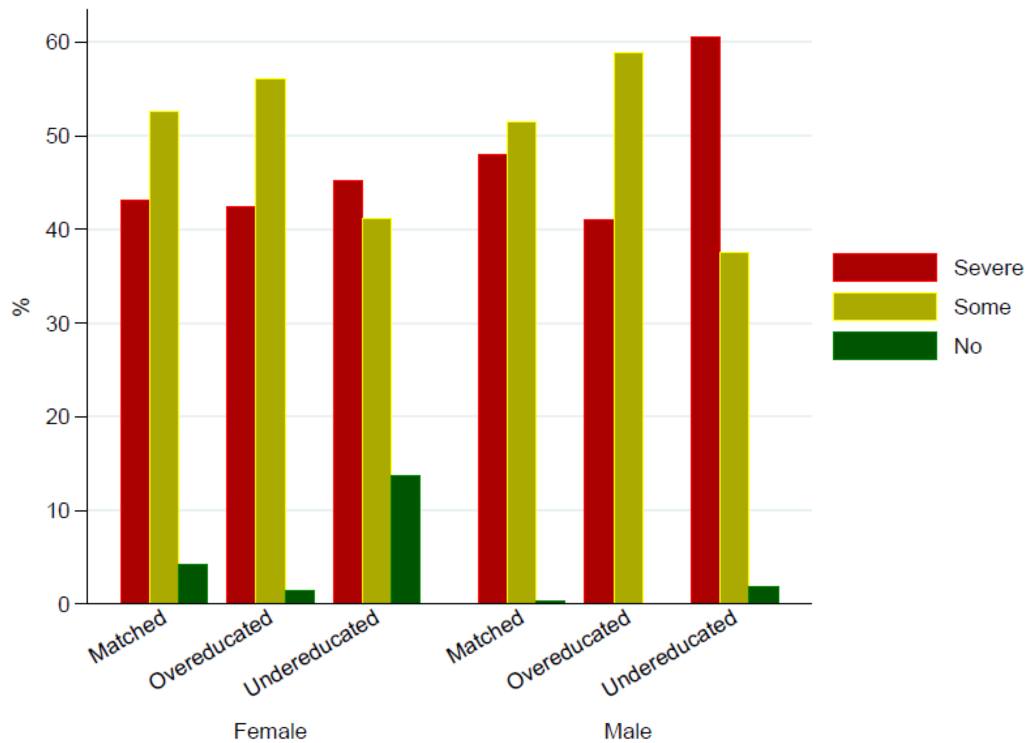
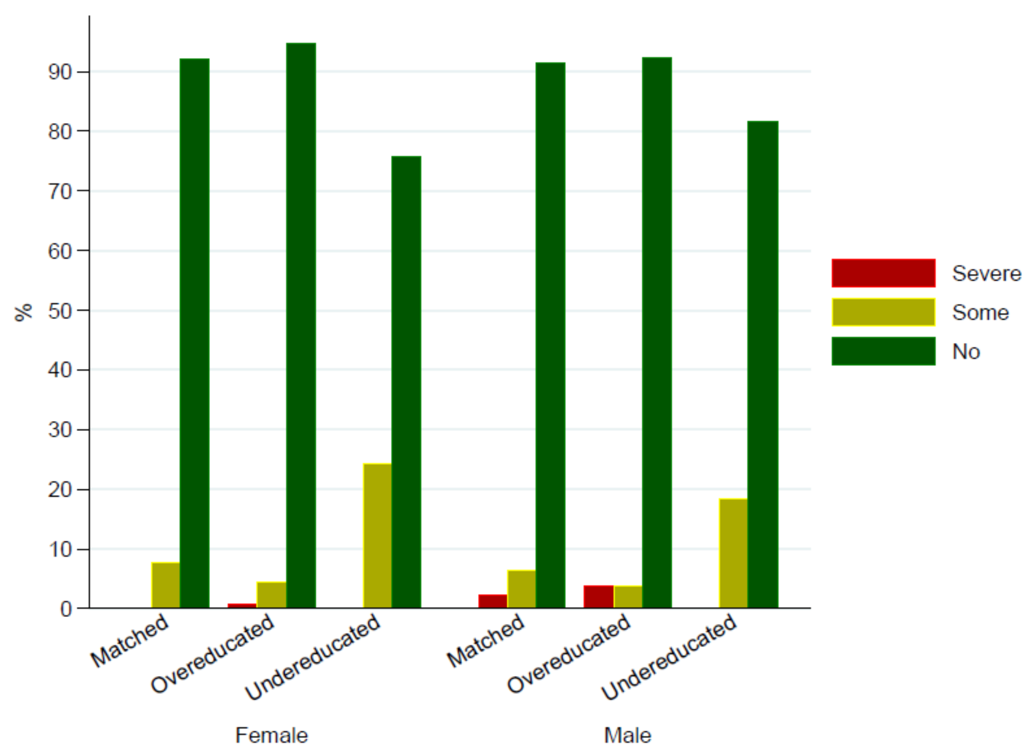


Figure A3 - *Distribution of vignettes responses over educational mismatch and gender: Depression*

Vignette 1:



Vignette 2:



Vignette 3:

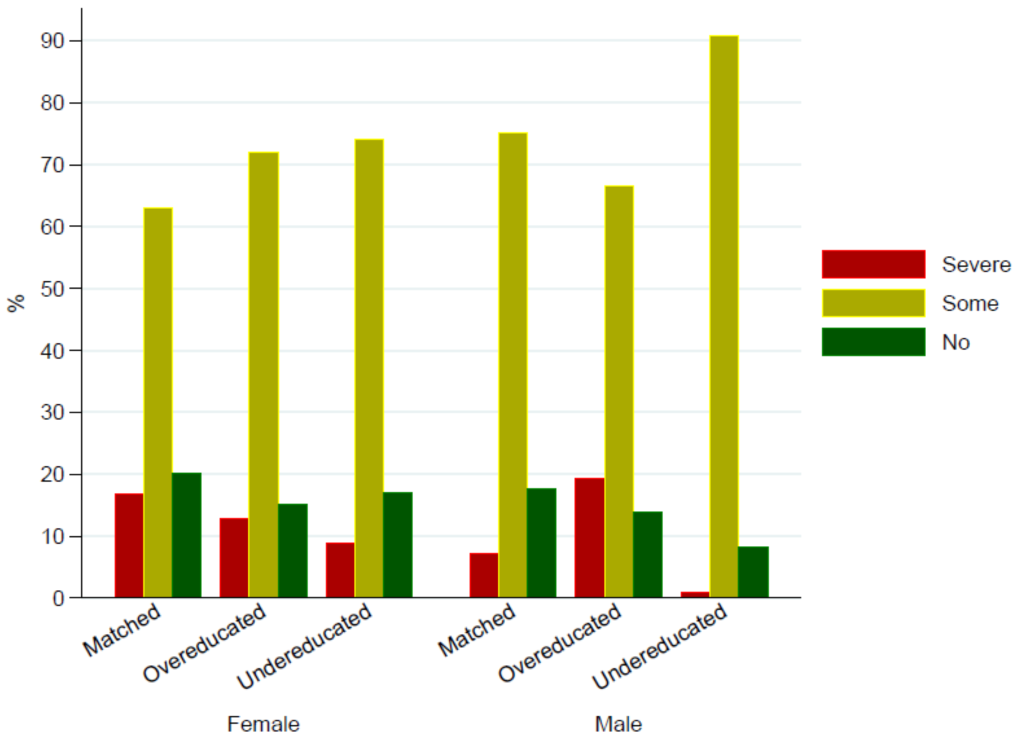


Table A3 – Descriptive statistics: t-test of equal means

	RLMS-HSE (2005) (n=2,973)		Sample with vignettes (n=2,133)			Weighted sample with vignettes (n=2,133)		
Variables	Mean	St. Dev.	Mean	St. Dev.	P-value	Mean	St. Dev.	P-value
Sociodemographic controls:								
Age	37.717	9.552	37.915	9.674	0.259	38.002	9.626	0.104
Gender	0.447	0.497	0.461	0.499	0.128	0.451	0.498	0.655
Single	0.264	0.441	0.213	0.410	0.000***	0.247	0.431	0.036**
12+ years of education	0.583	0.493	0.678	0.467	0.000***	0.580	0.494	0.706
Income (ln)	8.622	0.718	8.665	0.708	0.001***	8.626	0.707	0.790
Work-related controls:								
Part-time work	0.078	0.268	0.079	0.270	0.760	0.076	0.265	0.760
Over-time work	0.404	0.491	0.390	0.488	0.106	0.401	0.490	0.694
White collar	0.560	0.496	0.624	0.485	0.000***	0.567	0.496	0.460
Main explanatory variables:								
Over-education	0.155	0.362	0.174	0.379	0.005***	0.155	0.362	0.985
Under-education	0.135	0.341	0.109	0.312	0.000***	0.141	0.348	0.281

**p<0.10, **p<0.05, ***p<0.01*