Erasmus
School of
Economics

Overschooling and Underschooling: Incidence and Wage Effects in the Netherlands

---

N.D. Janssen 416854

Bachelor Thesis

Supervisor: Peter Lapperre, MBA

Erasmus University Rotterdam

Erasmus School of Economics

Capacity group Business Economics

Section Finance

Erasmus

Abstract

This paper attempts to contribute to the economic overschooling literature by providing empirical evidence concerning the incidence and wage effects of over- and underschooling for the Netherlands in 2014. We use the most recent data from the Dutch Labor Supply Panel Database, constructed by the Netherlands Institute of Social Research. Three methods are employed for the measurements of over- and underschooling: self-assessment, job analysis, and realized matches. Cross-sectional regression results indicate a significant negative wage effect for overschooled individuals compared to their well-matched peers with the same educational attainment. Conversely, underschooled individuals earn significantly more than their well-matched peers, except for the regressions using the self-assessed measure of underschooling.

Keywords: *Human Capital, Return on Education, Overschooling, Underschooling, Wage Effects*

Inhoudsopgave

1. Introduction

1.1 Subject

Overschooling and underschooling refer to the respective positive and negative discrepancy between the education level of an individual and his or her educational job requirements. The aim of this paper is to determine (a) the incidence and magnitude of over- and underschooling, and (b) the related cross-sectional wage effects of over- and underschooling in the Netherlands using the rational choice framework to human capital investments initially developed by Mincer (1958), Schultz (1959, 1961), and Becker (1962).

1.2 Relevance

The incidence of over- or underschooling arguably proves a mismatch between workforce participants and occupations. In other words, an inefficient labor market. It is therefore relevant for both (public) policy and personal financial planning to understand the effects of inefficient labor allocation and the possibility of profitable human capital investments to improve it. With this paper we hope to contribute to the scientific literature by determining over- and underschooling and the wage effects of these phenomena using the latest (2014) data from the Labor Supply Panel Database composed by the Netherlands Institute of Social Research.

1.3 Research Questions

Various methods have been employed historically to determine the incidence and magnitude of over- and underschooling. The methods differ in their use of proxies for the educational job requirements. Leuven and Oosterbeek (2011) distinguish between three approaches: self-assessment, job analysis, and realized matches. The first approach is employed by Duncan and Hoffman (1981) for the United States and by Hartog and Oosterbeek (1988) for the Netherlands. The second approach is described by Eckaus (1964). However it has been criticized heavily, i.a. by Halaby (1994) and Hartog (2000), due to the costly nature of this approach and the reliability of such measures for economic purposes. The third approach is used by Verdugo and Verdugo (1989), who define adequate schooling as having an education level within one standard deviation of the average education level for a particular job, and Kiker et al. (1997), who define the job requirement level as the mode of completed education levels of individuals in a particular occupation. Since the methods to determine the incidence of over- and underschooling presented in the literature do not point in one particular correct direction, we have formulated the research question concerning the first part of this paper as follows:

"What is the incidence of over- and underschooling in the Netherlands in 2014 using the three different approaches, and how do the results differ?"

Furthermore, we determine the wage effects of over- and underschooling in the cross-section using the Mincer (1974) human capital earnings equation, extended by Verdugo and Verdugo (1989). The

Mincer equation is considered the standard model for (individual) earnings research (Chiswick, 1997). We will estimate the model three times, once for each measure of over- and underschooling. This leads us to our second research question:

"What are the signs and magnitudes of the wage effects of over- and underschooling in the Netherlands in 2014 using the (extended) Mincer model and the three over- and underschooling measures?"

1.4 Structure

The paper is structured as follows: Section II describes the theoretical framework, both for determining the incidence of over- and underschooling and the determination of the wage effect. Section III describes how data have been altered and includes the used (constructed) variables and descriptions. Section IV describes the methodology, both for the determination of over- and underschooling and the wage effect determination. In Section V the main findings of this paper are presented, as well as an interpretation of them. The limitations of this paper's research are discussed in Section VI. Section VII describes the conclusion and provides answers to the research questions.

2. Theoretical Framework

2.1 Measuring Over- and Underschooling

The incidence of over- and underschooling is typically determined by comparing an individuals educational level and the educational requirements of the job in question. While an individual's educational level can often be observed directly, a proxy has to be employed for the educational job requirements. Leuven and Oosterbeek (2011) distinguish between three approaches to determine these requirements. Each approach has its disadvantages in the form of some degree of measurement error. We will discuss the approaches and relevant corresponding overeducation literature in the following sections.

2.1.1 Self-Assessment

The self-assessment approach is based on directly asking an individual how well his or her educational qualifications correspond to the job requirements in question. Formulation of the question is crucial to get the most accurate answers, that is, answers corresponding to the object of research: over- and underschooling. Perhaps unsurspringly, the question is formulated differently across studies as different authors have diverse interpretations of job requirements (Leuven and Oosterbeek, 2011). Questions differ in their emphasis on formal schooling (Duncan and Hoffman, 1981), informal schooling (Galasi, 2008), necessary (educational) preparation (Hartog and Oosterbeek, 1988), and (educational) performance requirements (Alba-Ramirez, 1993).

The direct approach of self-assessment has (in principle) the advantage of complete information: who knows a job's requirements better than the person who does it every day? Potential disadvantages include the possibility of behavioral biases due to the subjective nature of the approach and the fact that an individual is only completely informed regarding his or her own educational level and job requirements.

2.1.2 Job Analysis

The job analysis approach is based on occupational classifications constructed by analysts for each job. An example of an occupational classification index in the Netherlands is the *Standaard Beroepen Classificatie (SBC)*, which classifies each occupation by job kind and educational level. The education level ranges from Elementary (1) to Scientific (5)[1]. Kiker et al. (1997) employed this approach in the context of over- and underschooling wage effects in Portugal, using a Portugese classification.

Measuring job requirements through occupational classifications has the advantage of no individual behavioral biases due to the objective nature of the classification. However, the abstraction and standardisation of occupations does not allow for a range of requirements as observed in practice. Furthermore, classifying all occupations is a costly process, hence updates are often infrequent.

2.1.3 Realized Matches

The realized matches approach is a data based approach: job requirements are defined as some statistic relation between jobs and the educational level of employees in that job. Verdugo and Verdugo (1989) consider an individual adequately schooled (i.e. education level matches job requirements) if the individual's education level is within one standard deviation of the mean of educational levels within his or her occupation. Kiker et al. (1997) proposed using the mode of educational levels within an occupation as the definition of the educational job requirement. Groeneveld and Hartog (2003) use the educational requirements posted in job offers of the personnel department of a firm. Gottschalk and Hansen (2003) define over- and underschooling as having college or non-college jobs based on college premiums.

The realized matches approach results in an objective measure of job requirements, hence individual behavioral biases are diminished. Disadvantageous is the arbitrariness of the statistical relation used to define over- and underschooling and the lack of support for interoccupational variation. Furthermore, the college premium approach has limited relevance for non-U.S. research.

[1] Full SBC specification (Dutch): http://tinyurl.com/y867nwjw

2.2 Determining Wage Effects

2.2.1 Mincer Model

The model and earnings equation developed by Mincer (1958, 1974) has been a "cornerstone of empirical economics" (Heckman et al., 2003, p. 1) with applications in a range of subjects, including returns to schooling estimations and the male-female wage gap. The wide application of the equation developed by Mincer can be attributed by its simplicity and empirical success. The basic (OLS) Mincer equation takes the form of Formula (1).

(1) $$\ln w_i = a_0 + \rho_s s_i + \beta_0 x_i + \beta_1 x_i^2 + \beta_2 t_i + \beta_3 \ln\ W_i + \varepsilon_i$$

Here, wage is a function of years of schooling s, years of work experience x, years of tenure t, and the natural logarithm of hours worked W. If Mincer's assumptions are met, ρ_s denotes the return on the years of schooling. Conform the standard OLS assumptions, the expected value of the error term is zero: $E(\varepsilon) = 0$. Duncan and Hoffman (1981) and Verdugo and Verdugo (1989) use modified versions of the Mincer formula to estimate the wage effects of over- and underschooling. The modifications made by these authors accord to two methods that can be observed in the scientific literature to determine the (un)profitability of over- and underschooling in a cross-sectional setting: return on years of schooling and wage effects using dummies. We will elaborate on both in the following sections.

2.2.2 Duncan and Hoffman Model

Duncan and Hoffman (1981) measure over- and underschooling in years, allowing them to calculate the return on an additional year of over- or underschooling. The original Mincer equation is modelled for calculating the return to years of schooling in general and does not distinguish between adequate, over- and underschooling. To facilitate their research, Duncan and Hoffman modified the Mincer equation. Let Formula (2) denote the (abbreviated) basic Mincer equation.

(2) $$\ln w_i = a_0 + \delta_a S_i^a + \beta x_i + \varepsilon_i$$

Here, wage *w* is a function of schooling S_i^a and a vector of other variables x_i with coefficienten vector β. Now, by replacing S_i^a with the identity of Formula (3), Duncan and Hoffman allow for different effects for adequate, over- and underschooling years, as displayed in Formula (4).

(3) $$S_i^a \equiv S_i^r + S_i^u + S_i^o$$

(4) $$\ln w_i = a_0 + \delta_r S_i^r + \delta_u S_i^u + \delta_o S_i^o + \beta x_i + \varepsilon_i$$

Here, δ_r, δ_o, and δ_u denote the respective coefficients of required (S_i^r), over- (S_i^o) and underschooled (S_i^u) years. The main disadvantage is similar to one of the disadvantages of the Mincer model: schooling is not always presented in years in the data.

2.2.3 Verdugo and Verdugo Model

The data problem posed by both the previous models is non-existent in the Verdugo and Verdugo (1989) model. Instead of using years of schooling as independent over- or underschooling variable, they calculate dummies. The regression formula employed by Verdugo and Verdugo is displayed as Formula (5).

$$\ln w = a_0 + \delta_u \emptyset_i^u + \delta_o \emptyset_i^o + \beta x_i + \varepsilon_i \quad (5)$$

Here, x_i is a vector of other variables with coefficientvector β. $\emptyset_i^u$ and $\emptyset_i^o$ represent the under- and overschooling dummy, respectively[2]. Due to the fact that the under- and overschooling variables are no longer continuous, it is no longer possible to estimate the returns to individual years of schooling. However, the wage effects of a mismatch between educational level and job requirements can still be determined[3].

3. Data

3.1 Source

To investigate the incidence and magnitude of over- or underschooling and the wage effects of these phenomena, we use the latest (2014) data from the Labor Supply Panel Database (LSPD) composed by the Netherlands Institute for Social Research (Dutch: Sociaal Cultureel Planbureau). The database is part of the national project Supply of Labour and was founded in order to map the various aspects of the labor supply force in the Netherlands. Themes include labor mobility and job search behaviour, and variables such as tenure and job opinions. The database is based on a biannual survey filled out by a representative sample of Dutch households. The variables used in the determination of over- and underschooling and the regressions are either taken directly from the database, or constructed from the variables available in the database.

3.2 Manipulations

Not all data available was relevant or useful for our research purposes. Several observations have been dropped. The resulting database contained observations for wage labourers and independent contractors without a second or third job. Individual workers included civil servants, employees and business owners. Furthermore, observations were dropped for individuals that were unable to answer the questions concerning (a) the relation between their educational level and their job requirements, (b) work experience, (c) earnings, (d) education level, and (e) the year they started their current job.

[2] Naturally, there is no dummy for adequate schooling; this would result in a dummy variable trap.

[3] The signs of the coefficients in the Duncan and Hoffman model and the Verdugo and Verdugo are necessarily opposite, due to the fact that they estimate different things. The former concerns the yearly return to schooling. The latter determines the effect of being over- or underschooled on the wage rate.

Additionally, any non-relevant variables were dropped from the database. The resulting variables and the constructed variables are described in Table 1.

TABLE 1: *Descriptions of the variables used.*

Variables	Description
OVERSCHOOLING	Dummy; 1= Overschooled, 0 = Not Overschooled
UNDERSCHOOLING	Dummy; 1 = Underschooled, 0 = Not Underschooled
SEX	Sex; 1 = male, 2 = female
TENURE	Constructed variable: 2014 minus the year one started working at his or her current job.
TENURE_SQ	Constructed variable: TENURE squared.
WORK_EXPERIENCE	Number of years of experience in paid work; max. 51
WORK_EXPERIENCE_SQ	Constructed variable: WORK_EXPERIENCE squared.
lnEARNINGS_MONTHLY	Constructed variable: natural logarithm of monthly earnings.
lnHOURS_WEEK_ACTUAL	Constructed variable: natural logarithm of actual weekly hours worked.
EDUCATION_LEVEL	Highest degree in terms of level, according to the SOI-2006 first digit; 2 = elementary, 3 = VMBO/LBO/MAVO, 4 = MBO/HAVO/VWO, 5 = HBO, 6, 7 = WO.
EMPLOYMENT_TYPE	Type of employment; 1 = civil servant, 2 = employee, 4 = CEO, 5 = business owner.
EDUC_TEN	Interaction variable; EDUCATION_LEVEL multiplied by TENURE.
EDUC_EXP	Interaction variable; EDUCATION_LEVEL multiplied by EXPERIENCE.
SECTOR	Sector; 0 = Agriculture, 1 = Industrial, 2 = Construction, 3 = Commerce, Catering & Repair, 4 = Transport, 5 = Business Services, 6 = Care, 7 = Other Services, 8 = Government, 9 = Education.
FIRM_SIZE_CAT	Constructed categorical variable from continuous firm size variable; 0 = 0 – 5 employees, 1 = 5 – 50, 2 = 50 – 100, 3 = 100 – 200, 4 = 200 – 500, 5 = 500 – 1000, 6 = 1000+.
REGION	Province one lives; 1 = Groningen, 2 = Friesland, 3 = Drenthe, 4 = Overijssel, 5 = Flevoland, 6 = Gelderland, 7 = Utrecht, 8 = Noord-Holland, 9 = Zuid-Holland, 11 = Zeeland.

3.3 Over- and Underschooling

Next, we calculate over- and underschooling using the obtained data. Leuven and Oosterbeek (2011) distinguished between three different methods. In order to get a complete and accurate picture of over- and underschooling in the Netherlands, we will use all three methods as far as the data allows it. To account for any differences in sex, we have calculated the three measures of over- and underschooling thrice: once for the full sample, once for males, and once for females.

The self-assessed or perceived measure is based on a variable contained in the database. It is based on the question: "What is the relation between your education and knowledge and the work you perform?". If the answer was: "Lower than the work requires", we defined it as underschooling. Vice-versa, if the answer was: "Higher than the work requires", we defined it as overschooling.

The job analysis measure is based on the educational job requirement as per SBC 2010 classification. Overschooling defined as having an education level that is higher than the SBC job requirement. Conversely, underschooling is defined as having an educational level lower than the SBC job requirement.

Finally, our realized matches approach follows that of Kiker et al. (1997): for each job the modal educational level is calculated. Then, over- and underschooling are defined as having an education level above or below this mode, respectively. The results are summarized in Table 2.

TABLE 2: *The three measures of over- and underschooling for the full sample, for males exclusively, and for females exclusively.*

Description	Full Sample			Male			Female		
	Perc.	SBC	Mode	Perc.	SBC	Mode	Perc.	SBC	Mode
Completed = Required	64.15	62.75	68.99	62.58	60.58	68.56	65.75	64.97	69.44
Completed > Required	32.98	19.46	20.85	34.36	18.33	22.09	31.58	20.61	19.59
Completed < Required	2,87	17.79	10.16	3.07	21.09	9.36	2.66	14.42	10.97
Total	*100.00*	*100.00*	*100.00*	*100.00*	*100.00*	*100.00*	*100.00*	*100.00*	*100.00*

Striking is the very low occurence of underschooling in the perceived measure compared to the objective measures. The low occurance of underschooling using the perceived measure might be attributed to *overconfidence bias*: the phenomenon that people often overestimate their own qualities, i.a. described by Svensson (1981) in the context of self-assessment of driving skills. The low

incidence of underschooling can also be attributed to the question itself: not just formal education, but also knowledge is adressed. This can be interpreted in various ways. We will elaborate on this (significant) limitation in Section 6.

4. Methodology

4.1 Regression Model

We test the various measures of over- and underschooling separately in a cross-sectional OLS setting, resulting in three different tables of results. The framework used is based on the Verdugo and Verdugo (1989) model that extends the standard Mincer regression equation. This means that over- and underschooling will be measured in dummy terms, in contrast with over- and underschooling in years as pertained in Duncan and Hoffman (1981). Conform both frameworks, control variables are added, resulting in Formula (6).

$$\ln w = a_0 + \delta_u \emptyset_i^u + \delta_o \emptyset_i^o + \beta x_i + \psi \kappa_i + \varepsilon_i \tag{6}$$

Here, $\emptyset_i^o$ and $\emptyset_i^u$ denote the over- and underschooling dummies, respectively. x_i is a vector of variables: work experience (squared), tenure (squared), employment type, sex, log of hours worked weekly, and interaction terms for education and tenure, and education and work experience. κ_i is a vector of control variables: education level, region, sector, and firm size. Both β and ψ are vectors containing coefficients.

Before we add our variables of interest, we first estimate the basic model. The results of which are displayed in Appendix Table A. In the regression results following the basic model, we will omit the control variables for clarity. Furthermore, to avoid any biases due to sex, three regressions have been estimated for the basic model and each measure of over- and underschooling: one for the full sample, one for males exclusively, and one for females exclusively

5. Results

5.1 Perceived Measure

Firstly, we estimate the regression model extended with the perceived (self-assessment) measure of over- and underschooling. The results of which are presented in Table 3. The R^2 is relatively high: 55.9% of the total variation is explained by the model for males, whereas 80.6% is explained for females. The full sample has an R^2 of 77.1%.

The overschooling coefficient is significant on the 0.1% level for both male and female workers and has a negative sign. The underschooling coefficient is insignificant. This can be interpreted as follows: individuals that work in an occupation with lower educational requirements than their educational level, earn less than individuals with similar educational attainment that have an occupation which

has the same educational requirements as their educational level. The underschooling coefficient is insignificant. This is consistent with the abnormally low occurence of underschooling in the perceived measure (see Table 2).

TABLE 3: *Regression estimations of the base model extended with perceived measures of over- and underschooling.*

Variables	Full Sample	Male	Female
	β (t-stat)	*β (t-stat)*	*β (t-stat)*
SEX	-0.0723***	-	-
	(-5.54)		
OVERSCHOOLING	-0.0731***	-0.0542***	-0.103***
	(-6.41)	(-3.51)	(-6.29)
UNDERSCHOOLING	-0.00258	0.0145	-0.0206
	(-0.08)	(0.34)	(-0.44)
EMPLOYMENT_TYPE	-0.0109	-0.0196	-0.00741
	(-0.47)	(-0.57)	(-0.24)
WORK_EXPERIENCE	0.0243***	0.0293***	0.0246***
	(6.32)	(5.72)	(4.24)
WORK_EXPERIENCE_SQ	-0.000295***	-0.000470***	-0.000200**
	(-7.02)	(-8.45)	(-3.13)
TENURE	0.00883*	0.00690	0.0121*
	(2.49)	(1.60)	(2.04)
TENURE_SQ	-0.0000690	0.0000426	-0.000231**
	(-1.41)	(0.69)	(-2.99)
lnHOURS_WEEK_ACTUAL	0.840***	0.592***	0.906***
	(55.76)	(19.16)	(51.15)
EDUC_TEN	-0.000841	-0.00136	-0.000257
	(-1.27)	(-1.71)	(-0.23)
EDUC_EXP	-0.000871	0.000249	-0.00235*
	(-1.35)	(0.30)	(-2.37)
CONSTANT	4.149***	4.922***	4.052***
	(34.72)	(27.70)	(20.71)
N	2364	1188	1176
R^2	0.771	0.559	0.806

Notes: T-statistics in parentheses. Coefficients are statistically insignificant unless indicated as follows: * (5% level), ** (1% level), *** (0.1% level). Control variables are not shown in the table. However, they have been included.

Work experience and its squared term are significant for both male and female workers, with a positive and (small) negative coefficient, respectively. This indicates diminishing wage increases as workers gain more work experience. A similar effect is observed for tenure, although only significant in the female cohort. The log of the weekly hours worked is (naturally) significant and positive for all

cohorts. Finally, the interaction term education times experience is slightly negative and significant for the female cohort, indicating a lower work experience effect for higher educated workers.

5.2 SBC Measure

Secondly, we estimate the regression model extended with the objective SBC (job analysis) measure of over- and underschooling. The results of which are presented in Table 4. Again, the R^2 of the model is quite high: 81.2% of the total variation in the dependent variable are explained by the model for females, whereas the R^2 for the male cohort is 58.4%. The full sample R^2 is 78.1%.

TABLE 4: *Regression estimations of the base model extended with SBC measures of over- and underschooling.*

Variables	Full Sample	Male	Female
	β (t-stat)	*β (t-stat)*	*β (t-stat)*
SEX	-0.0724***	-	-
	(-5.67)		
OVERSCHOOLING	-0.123***	-0.114***	-0.136***
	(-8.53)	(-5.84)	(-6.59)
UNDERSCHOOLING	0.114***	0.125***	0.115***
	(7.20)	(6.17)	(4.76)
EMPLOYMENT_TYPE	-0.0142	-0.0205	-0.0121
	(-0.62)	(-0.62)	(-0.40)
WORK_EXPERIENCE	0.0197***	0.0242***	0.0202***
	(5.20)	(4.84)	(3.52)
WORK_EXPERIENCE_SQ	-0.000251***	-0.000419***	-0.000161*
	(-6.08)	(-7.73)	(-2.55)
TENURE	0.00777*	0.00549	0.0119*
	(2.24)	(1.31)	(2.03)
TENURE_SQ	-0.0000682	0.0000431	-0.000242**
	(-1.43)	(0.72)	(-3.17)
lnHOURS_WEEK_ACTUAL	0.812***	0.550***	0.880***
	(53.96)	(18.10)	(49.16)
EDUC_TEN	-0.000649	-0.00106	-0.000209
	(-1.00)	(-1.36)	(-0.19)
EDUC_EXP	-0.000397	0.000649	-0.00177
	(-0.63)	(0.79)	(-1.81)
CONSTANT	4.279***	5.079***	4.176***
	(36.20)	(29.15)	(21.50)
N	2364	1188	1176
R^2	0.781	0.584	0.812

Notes: T-statistics in parentheses. Coefficients are statistically insignificant unless indicated as follows: * (5% level), ** (1% level), *** (0.1% level). Control variables are not shown in the table. However, they have been included.

Here, both the coefficients for over- and underschooling are significant on the 0.1% level. Conform the perceived measure, the overschooling sign is negative. Conversely, the underschooling sign is positive. Following the previous interpretation of overschooling, the positive underschooling signs can be explained in terms of education and job requirement match: individuals working above their educational level earn more than their counterparts with education levels matching their job requirements.

Coefficients and their signs of work experience (squared), tenure (squared), and the natural logarithm of weekly hours worked are similar to the estimations for the perceived over- and underschooling measure. However, the significant effect of the (female) interaction term of education level multiplied by work experience has vanished.

5.3 Modal Measure

Finally, we estimate the regression model using the model measure of over- and underschooling. The regression results are presented in Table 5. Similar to the previous models, the model has a relatively high explanatory power, with an R^2 of 80.3% for females and 56.3% for males. The full sample has an R^2 of 77.1%.

Using the modal measure, all over- and underschooling coefficients are significant on the 0.1% level except for male overschooling and female underschooling. The latter two coefficients are significant on the 5% level. The signs of the relevant coefficients are similar to those found in the model based on the SBC measure: positive for underschooling and negative for overschooling. The interpretation of the coefficients is therefore similar the previous models: the mismatch between job requirements and educational level has a positive underschooling effect compared to well-matched peers and vice versa.

Although some coefficients show similar significance and sign, others stand out. Work experience (squared), female tenure (squared), weekly hours worked, and the education and work experience interaction term are similar to the SBC measure model. However, in this model, tenure is also significant and postive for males, indicating a beneficial effect as one works longer for the same employer.

TABLE 5: *Regression estimations of the base model extended with the modal measure of over- and underschooling.*

Variables	Full Sample	Male	Female
	β (t-stat)	*β (t-stat)*	*β (t-stat)*
SEX	-0.0673***	-	-
	(-5.17)		
OVERSCHOOLING	-0.0601***	-0.0366*	-0.0823***
	(-4.50)	(-2.09)	(-4.08)
UNDERSCHOOLING	0.0780***	0.108***	0.0551*
	(3.99)	(3.96)	(2.00)
EMPLOYMENT_TYPE	-0.0120	-0.0211	-0.00910
	(-0.52)	(-0.62)	(-0.29)
WORK_EXPERIENCE	0.0235***	0.0284***	0.0228***
	(6.11)	(5.58)	(3.90)
WORK_EXPERIENCE_SQ	-0.000284***	-0.000460***	-0.000184**
	(-6.77)	(-8.32)	(-2.85)
TENURE	0.00901*	0.00622	0.0150*
	(2.54)	(1.45)	(2.49)
TENURE_SQ	-0.0000777	0.0000299	-0.000250**
	(-1.59)	(0.49)	(-3.21)
lnHOURS_WEEK_ACTUAL	0.845***	0.603***	0.906***
	(56.21)	(19.72)	(50.64)
EDUC_TEN	-0.000806	-0.00112	-0.000719
	(-1.22)	(-1.40)	(-0.63)
EDUC_EXP	-0.000794	0.000330	-0.00209*
	(-1.23)	(0.40)	(-2.09)
CONSTANT	4.072***	4.822***	3.944***
	(33.99)	(27.22)	(20.06)
N	2364	1188	1176
R^2	0.771	0.563	0.803

Notes: T-statistics in parentheses. Coefficients are statistically insignificant unless indicated as follows: * (5% level), ** (1% level), *** (0.1% level). Control variables are not shown in the table. However, they have been included.

6. Limitations and Extensions

6.1 Measurement Error

The three methods employed to measure over- and underschooling in the sample all have their advantages and drawbacks. The self-assessment methods is on paper one of the best methods, as workers themselves have the most informed concerning their job requirements. However, the subjective nature of this method might result in behavioral biases. It is unclear whether this paper's relatively low incidence of (self-assessed) underschooling is due to some behavioral bias such as the overconfidence effect, or if it is the result of a to broadly formulated question.

Objective measures such as the SBC based method and the mode approach can be assumed to be objective measures of the labor mismatch. This does not mean that these methods are exempt from criticism: the statistical relation to define the mismatch is arbitrary at best. The SBC method employed here also does not use some (standard deviation based) range for adequate schooling as a result of data constraint: the variable used to measure individual education level is limited to a five level scale.

6.2 Omitted Variable Bias

Another issue affecting the validity of the obtained results is omitted variable bias. If a variable exists that has an effect on both the independent variable (log earnings) and the dependent variables, then the obtained coefficient estimates are biased. A possible extension of this paper would be the use of instrumental variables, e.g. Korpi and Tahlin (2009). However, finding proper instrumental variables is hard, if not impossible. Another possibility is the use of a fixed-effects regression model (cf. Leuven and Oosterbeek, 2011).

6.3 Wage Effect Model Limitations

Wage effects of over- and underschooling have been measures solely using the Verdugo and Verdugo (1989) framework. This dummy approach of over- and underschooling does not allow an interpretation of the relevant coefficients in terms of a return to (over- or under)schooling per year. This important limitations is mainly due to data constraints: the variable used to measure individual education level is limited to a five level scale. Although also categorical, the database does contain a more extensive 'completed education' variable. However, the levels are not transitive: some types of education overlap. An extension of this paper could be the translation of the values of this variable into years of schooling, allowing for the Duncan and Hoffman (1981) framework to be usable.

7. Conclusion

Now that we have the results and have assessed the limitations of these results, we are able to formulate the answers to the two research questions. Starting with the first question:

"What is the incidence of over- and underschooling in the Netherlands in 2014 using the three different approaches, and how do the results differ?"

The three measures of over- and underschooling produce strikingly different results. The incidence of overschooling using the self-assessment method is approximately 33% across the full sample, males, and females. The incidence of underschooling is very small compared to the other two measures: approximately 3%. The SBC method produces a fairly balanced result, with a slightly lower incidence of underschooling than overschooling: approximately 18% and 19%, respectively. The figures for the modal method are in between the two methods previously described: overschooling and underschooling average 21% and 10%, respectively.

The regression results using the three different measures of over- and underschooling indicate a significant positive effects for underschooling for both males and females, except for the perceived measure. Conversely, we find a significant negative effect for overschooled workers for all measures. The significant variables for over- and underschooling can be interpreted in terms of the match between job requirements and educational levels of an individual: overschooled workers earn less than their well-matched peers with equal educational attainment. Vice-versa, underschooled workers earn more than their well-matched peers. This allows us to answer the second research question:

"What are the signs and magnitudes of the wage effects of over- and underschooling in the Netherlands in 2014 using the (extended) Mincer model and the three over- and underschooling measures?"

Where significant, the signs of the underschooling coefficients are positive. The overschooling coefficients are significant and positive for all three measures. We are cautious to interpret the magnitude of the coefficients of the regression results due to (a) measurement error in the over- and underschooling variables, and (b) possible omitted variable bias in the regression. That said, the overschooling coefficient ranges are summarized in Table 6.

TABLE 6: *Ranges of the over- and underschooling coefficients.*

Description	Full Sample		Male		Female	
	Min.	Max.	Min.	Max.	Min.	Max.
Overschooling	-0.0601	-0.123	-0.0366	-0.114	-0.0823	-0.136
Underschooling	0	0.114	0	0.125	0	0.115

Notes: the minimal values of zero are due to the insignificance of the coefficient for the basic model extended with the perceived measure.

Although overschooling has a negative coefficient across all measures, it is not necessarily the case that schooling beyond the adequate level has a negative return, nor that extra schooling beyond the adequate level is inadvisable. On the aggregate, overschooled workers earn less than their adequately schooled counterparts. However, it is quite probable that one year of overschooling has a considerably lower negative effect than five years of overschooling. Further research is necessary for concrete policy recommendations, both for employees and employers.

References

Alba-Ramirez, A. (1993). Mismatch in the Spanish labor market: overeducation?. *Journal of Human Resources*, 259-278.

Becker, G. S. (1962). Investment in human capital: A theoretical analysis. *Journal of political economy*, *70*(5, Part 2), 9-49.

Chiswick, B. R. (1997). Interpreting the coefficient of schooling in the human capital earnings function.

Duncan, G. J., & Hoffman, S. D. (1981). The incidence and wage effects of overeducation. *Economics of Education Review*, *1*(1), 75-86.

Eckaus, R. S. (1964). Economic criteria for education and training. *The Review of Economics and Statistics*, 181-190.

Galasi, P. (2008). *The effect of educational mismatch on wages for 25 countries* (No. BWP-2008/8). Budapest Working Papers on the Labour Market.

Gottschalk, P., & Hansen, M. (2003). Is the proportion of college workers in noncollege jobs increasing?. *Journal of Labor Economics*, *21*(2), 449-471.

Groeneveld, S., & Hartog, J. (2004). Overeducation, wages and promotions within the firm. *Labour Economics*, *11*(6), 701-714.

Halaby, C. N. (1994). Overeducation and skill mismatch. *Sociology of education*, 47-59.

Hartog, J. (2000). Over-education and earnings: where are we, where should we go?. *Economics of education review*, *19*(2), 131-147.

Hartog, J., & Oosterbeek, H. (1988). Education, allocation and earnings in the Netherlands: 0verschooling?. *Economics of Education Review*, *7*(2), 185-194.

Heckman, J. J., Lochner, L. J., & Todd, P. E. (2003). *Fifty years of Mincer earnings regressions* (No. w9732). National Bureau of Economic Research.

Kiker, B. F., Santos, M. C., & De Oliveira, M. M. (1997). Overeducation and undereducation: evidence for Portugal. *Economics of Education Review*, *16*(2), 111-125.

Korpi, T., & Tåhlin, M. (2009). Educational mismatch, wages, and wage growth: Overeducation in Sweden, 1974–2000. *Labour Economics*, *16*(2), 183-193.

Leuven, E., & Oosterbeek, H. (2011). Overeducation and mismatch in the labor market. *Handbook of the Economics of Education*, *4*, 283-326.

Mincer, J. (1958). Investment in human capital and personal income distribution. *Journal of political economy*, *66*(4), 281-302.

Mincer, J. A. (1974). The human capital earnings function. In *Schooling, Experience, and Earnings* (pp. 83-96). NBER.

Schultz, T. W. (1959). Investment in man: An economist's view. *Social Service Review*, *33*(2), 109-117.

Schultz, T. W. (1961). Investment in human capital. *The American economic review*, *51*(1), 1-17.

Svenson, O. (1981). Are we all less risky and more skillful than our fellow drivers?. *Acta psychologica*, *47*(2), 143-148.

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

Appendix

APPENDIX TABLE A: *Regression results of the basic model, estimated for the full sample, male exclusively, and female exclusively.*

Variables	Full Sample	Male	Female
	β (t-stat)	*β (t-stat)*	*β (t-stat)*
SEX	-0.0632***	-	-
	(-4.83)		
EMPLOYMENT_TYPE	-0.00954	-0.0150	-0.0101
	(-0.41)	(-0.44)	(-0.32)
WORK_EXPERIENCE	0.0190***	0.0297***	0.0116***
	(9.78)	(11.32)	(4.00)
WORK_EXPERIENCE_SQ	-0.000261***	-0.000456***	-0.000141*
	(-6.43)	(-8.55)	(-2.24)
TENURE	0.00560**	0.000985	0.0123***
	(3.14)	(0.43)	(4.51)
TENURE_SQ	-0.0000728	0.0000485	-0.000248**
	(-1.50)	(0.79)	(-3.21)
lnHOURS_WEEK_ACTUAL	0.853***	0.607***	0.918***
	(56.48)	(19.70)	(51.16)
EDUC_TEN	-.0008432	-.0013626*	-.000299
	(-1.26)	(-1.70)	(-0.26)
EDUC_EXP	-.0008503	.0002604	-.0022562**
	(-1.31)	(0.31)	(-2.24)
EDUCATION_LEVEL	0.0311	0.0446	0.0578
vmbo/lbo/mavo	(0.56)	(0.62)	(0.68)
EDUCATION_LEVEL	0.143**	0.150*	0.173*
mbo/havo/vwo	(2.61)	(2.11)	(2.07)
EDUCATION_LEVEL	0.334***	0.363***	0.333***
hbo	(6.05)	(5.06)	(3.97)
EDUCATION_LEVEL	0.496***	0.514***	0.505***
wo	(8.79)	(6.99)	(5.91)
REGION	0.0111	-0.0343	0.0246
Friesland	(0.27)	(-0.59)	(0.44)
REGION	0.0152	0.0346	-0.0166
Drenthe	(0.37)	(0.59)	(-0.29)
REGION	0.0326	0.0342	0.0209
Overijssel	(0.94)	(0.69)	(0.44)
REGION	0.0201	0.0228	0.00780
Flevoland	(0.51)	(0.40)	(0.14)
REGION	0.0172	0.0363	-0.00631
Gelderland	(0.51)	(0.75)	(-0.14)
REGION	0.0504	0.0461	0.0316
Utrecht	(1.43)	(0.91)	(0.66)
REGION	0.0538	0.0698	0.0132

Noord-Holland	(1.67)	(1.49)	(0.30)
REGION	0.0723*	0.0851	0.0520
Zuid-Holland	(2.28)	(1.86)	(1.21)
REGION	0.00260	0.0889	-0.0927
Zeeland	(0.06)	(1.58)	(-1.66)
REGION	0.0431	0.0718	0.00162
Noord-Brabant	(1.33)	(1.54)	(0.04)
REGION	0.0271	0.0364	0.00700
Limburg	(0.76)	(0.71)	(0.14)
FIRM_SIZE_CAT	0.0811**	0.0749	0.0641
5 – 100	(2.68)	(1.73)	(1.55)
FIRM_SIZE_CAT	0.127***	0.129**	0.108*
100 – 500	(4.06)	(2.90)	(2.50)
FIRM_SIZE_CAT	0.145***	0.146**	0.126**
500+	(4.54)	(3.22)	(2.87)
SECTOR	-0.204***	-0.199**	-0.361*
Industry	(-3.32)	(-3.07)	(-2.34)
SECTOR	-0.194**	-0.168*	-0.419*
Construction	(-3.00)	(-2.49)	(-2.45)
SECTOR	-0.290***	-0.257***	-0.472**
Trade, Catering, Repair	(-4.74)	(-3.91)	(-3.12)
SECTOR	-0.267***	-0.240***	-0.514***
Transport	(-4.21)	(-3.53)	(-3.30)
SECTOR	-0.178**	-0.164*	-0.353*
Business services	(-2.94)	(-2.54)	(-2.33)
SECTOR	-0.274***	-0.312***	-0.434**
Care	(-4.49)	(-4.57)	(-2.88)
SECTOR	-0.182**	-0.182*	-0.354*
Other services	(-2.81)	(-2.57)	(-2.29)
SECTOR	-0.221***	-0.251***	-0.381*
Government	(-3.40)	(-3.45)	(-2.45)
SECTOR	-0.356***	-0.394***	-0.506***
Education	(-5.64)	(-5.56)	(-3.32)
CONSTANT	4.148***	4.871***	4.056***
	(35.91)	(28.77)	(20.76)
N	2364	1188	1176
R^2	0.767	0.553	0.7995