Overeducation's Impact on Wages across Industries and Occupations

A Regression Analysis using PIAAC Data from the Netherlands

Master Thesis

Rens Jacobus Wouter van der Geest

TU Delft

Overeducation's Impact on Wages across Industries and Occupations

A Regression Analysis using PIAAC Data from the Netherlands

by

Rens Jacobus Wouter van der Geest

Student number: 4567463

Master thesis submitted to Delft University of Technology in partial fulfilment of the requirements for the degree of

Master of Science in Management of Technology
Faculty of Technology, Policy and Management

To be defended in public on August 21th, 2024

Thesis committee:

Chairperson	Dr. ing. V. E. Scholten	Delft Centre for Entrepreneurship
First Supervisor	Dr. E. Schröder	Economics of Technology and Innovation
Second Supervisor	Dr. ing. V. E. Scholten	Delft Centre for Entrepreneurship

TU Delft

EXECUTIVE SUMMARY

Labor is changing due to globalization, increased mobility of workers, demographic changes, and changes due to the level and amount of educational attainment. This evolving landscape causes the requirements for labor to shift as well. A workforce with a higher education is necessary. The labor market responds to this demand, also influenced by policies promoting higher education. This landscape highlights the importance of the alignment between the supply and demand of labor. Given the need for a more highly educated workforce and the ambition to develop a highly educated population, educational mismatch becomes a relevant issue. Overeducation refers to a form of educational mismatch where individuals possess higher education levels than required for their jobs. The incidence of overeducation is growing and it is found that overeducation is an enduring and growing problem. Overeducation can affect wages. The problem with overeducation and its effect on wage lies in the consequences of its effect as it can lead to income inequality, economic inefficiency, and less returns on education.

This research is conducted for policymakers, employers, employees, and educational institutions in the Netherlands, which want to understand and address the wage effect of overeducation. The wage effect of overeducation has not been examined specifically focusing on the Netherlands and the influence of occupation or industry type has not been included in studies. By examining the wage effects of overeducation across industries and occupations, the findings aim to provide insights, which can be used to create policies that mitigate the negative consequences, provide information for employers and employees, and guide educational planning to better align with labor market demands. By including the industry and occupation types in this study, insights on the areas where overeducation and its wage effect is prevalent are identified. Hence, the objective of this study is to investigate and analyze the effect of overeducation on the wage of workers across occupations and industries in the Netherlands. Based on the research objective, the main research question is formulated:

To what extent does overeducation's effect on the wage of a worker vary across occupational groups and industries in the Netherlands?

To answer the research question, a literature review was conducted. In the literature it is discussed how overeducation is defined, how it can be measured, different theories explaining the effect of overeducation on wage, and models that can be used to determine the wage based on education. Overeducation is identified as a situation where a worker's educational level exceeds the requirements of their job. The choice for a certain measuring is often depending on the data that is available. In studies, it is found that the assignment theory often holds true. Therefore, the assignment theory, which explains the quality of the match to be a determinant of wage and uses determinants of both the demand and supply side of the labor market, combined with the ORU model, is used to base the theoretical framework, conceptual model, and part of the hypotheses on. Overeducation indicates a poor quality match between the job and the worker. As a result, according to the theory, overeducated workers experience a wage penalty compared to properly matched workers with the same obtained education. And, overeducated people are expected to receive a wage premium compared to properly matched workers with the same required education. Furthermore, two hypotheses are developed covering the influence of the occupation type and the industry type.

The research employs a quantitative approach, utilizing variables from the Programme for the International Assessment of Adult Competencies (PIAAC) data collected in the Netherlands during 2011-2012. The conceptual model, based on the ORU (Overeducation, Required education, Undereducation) model, is used to examine the wage effects of overeducation. A regression analysis is conducted to explore the relationships between overeducation, wages, and the moderating effects of industry and occupation types and calculate the coefficients belonging to the hypotheses to test them.

This study found that the overeducation results in a wage premium compared to properly matched workers in similar jobs but a wage penalty when compared to workers with the same level of education but

they are properly matched. Thus, the assignment theory is also true in the Dutch labor market. Subsequently, it is verified if the theory holds robustly across industry and occupation types. This was often the case, but one industry and one occupation showed different wage effects, namely the construction industry and the high skilled blue collar workers. These groups are included as dummies combined with overeducation to act as an interaction term in the final regression model. The model with the dummy variables provided a better fit to the data, compared to models which did not use these dummies. When a worker is in the construction industry or is a high-skilled blue-collar worker, the wage penalty appears smaller and the wage premium larger. However, no inference can be made about the moderating effect as the coefficient belonging to the moderating terms are insignificant. This answers the main research question.

This study has filled several knowledge gaps by (1) examining the effect of overeducation in the Dutch labor market, (2) validating the assignment theory, and (3) providing new insights about how the effect of overeducation differs across certain industries and occupations. With this knowledge, individuals can better assess the value of their educational investments, while educational institutions and policymakers can create policies to mitigate the negative impacts of overeducation. First, policymakers now know that the construction industry and high skilled blue collar workers seem to experience different wage effects due to overeducation. Although the causes of this difference need further investigation, it provides insights that occupation and industries do seem to have an influence on the effect of overeducation on wage. Improving this knowledge can provide information that can be used to target problem areas in terms of industries or occupations. Furthermore, the validation of the assignment theory offers direct guidance for policymakers to influence the quality of the match based on education between the worker and a job. Policies can be designed to better align the obtained education of the worker with the required education of the worker. Additionally, certain industries or occupations and their required educational paths might be promoted to improve the match between workers and jobs. Future research should address the limitations of this study by using longitudinal data, exploring additional variables, and conducting comparative analyses across different countries to examine the wage effect of overeducation across industries and occupations.

ACKNOWLEDGMENTS

This thesis marks the end of my studies at the TU Delft. I first arrived in Delft in 2016 to start my bachelor in Mechanical Engineering. Now, 8 years later, I am graduating with a master in Management of Technology. During this period I have learned a lot, both personally and academically. I have overcome doubts about my study choices and discovered what I enjoy most and am most interested in. As a result, my time in Delft has equipped me with the necessary knowledge to start the next chapter of my life.

Next, I would like to thank some people who supported me during my studies in Delft. First of all, I would like to thank Enno Schröder for his support, guidance, and feedback as the first supervisor of this thesis. You were kind, understanding, and knowledgeable. Your guidance helped me to keep going when I was stuck and elevate this thesis. I have learned a lot from you. Additionally, I extend my gratitude to Victor Scholten, my second supervisor. Although we have not spoken often, your questions and feedback have helped me improve this thesis. Finally, I would like to thank my family and friends. They always have supported me and pushed me to keep going when times were difficult. I would not have been able to do this without all of your support.

Rens J. W. van der Geest
Delft, July 2024

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

Acronym	Definition
CLM	Classical Linear Model
HSBC	High Skilled Blue Collar
HSWC	High Skilled White Collar
ISCO	International Standard Classification of Occupation
ISIC	International Standard Industrial Classification of All Economic Activities
LSBC	Low Skilled Blue Collar
LSWC	Low Skilled White Collar
MOT	Management of Technology
OLS	Ordinary Least Squares
ORU	Overeducation, Required education, Undereducation
PIAAC	Programme for the International Assessment of Adult Competencies

1

INTRODUCTION

Labor is changing. Globalization led to increased competition, resulting in a need for lower labor costs or jobs will be outsourced to other countries. The mobility of workers has increased. Demographic changes, such as an aging population, population growth or a change in the level and amount of educational attainment, can influence the demand and supply for a specific education and/or skills in the labor market. For example, technological advancements can displace jobs due to automation, while creating demand for higher educated individuals. Due to these changes in recent years, the requirements for labor will shift as well. It is shown that the increasing share of knowledge work in labor results in a need for a workforce with more education, cognitive abilities, and interpersonal skills (Handel, 2020). The supply side of the labor market responds and more people get a higher education (Brunello & Wruuck, 2021). This trend is also shown in the Netherlands, where it is their ambition (and of the European Union) to evolve into a more highly educated population (van der Mooren & de Vries, 2022). Vandeplas and Thum-Thysen (2019) explain that countries with a growth in higher education often also show an increasing supply of jobs requiring higher education levels. In this evolving landscape of the demand and supply of labor, the alignment between the supply and demand of labor is as important as ever. But when imbalances occur, the mismatch between the workers and the requirements for a job has emerged as a challenge. In literature, various types of mismatches are mentioned and used. A study by Shahidan and Ismail (2021) provided an overview of the terms used for the different types of mismatches. Figure 1.1 shows the different types of mismatches. A mismatch that is particularly relevant these days is the mismatch based on education, due to the need for higher educated people and the ambition of governments to evolve into a more highly educated population.

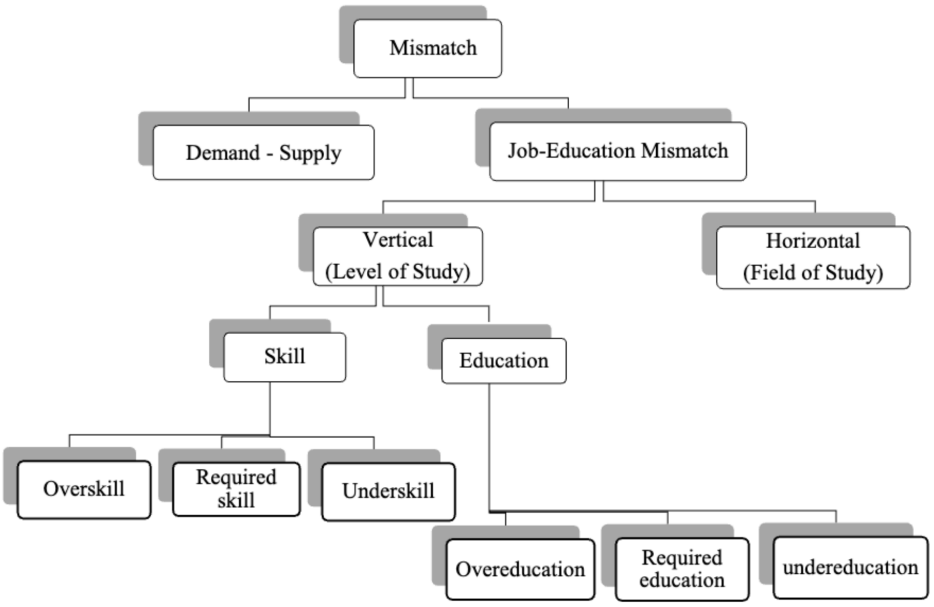

Figure 1.1.: An Overview of the Mismatch Types (Shahidan & Ismail, 2021)

The educational mismatch occurs when there is a discrepancy between the education required for a job and the education a worker has obtained. In this case, overeducation is defined as a situation where a worker that is employed has a higher degree than the job requires. For instance, a person with a master's degree in economics is employed as an administrative assistant. In this case, it is typically not necessary to have such a high degree. Logically, undereducation describes the opposite situation where an employee has a degree that is lower than the degree required in order to efficiently perform a job's responsibilities. A properly matched worker, on the other hand, has exactly the education required for their role, thus no educational mismatch occurs.

1.1. PROBLEM STATEMENT

The occurrence of mismatches poses a challenge with a negative impact for individuals, firms, and at the macro level (McGowan & Andrews, 2015; McGuinness, Pouliakas, & Redmond, 2018; Quintini, 2011). For individuals, it can have consequences for their wage and job satisfaction. Furthermore, mismatch can also have a negative impact on employers. For example, the horizontal mismatch can lead to workers not having the necessary knowledge of the field they are working in to efficiently complete their tasks, or complete them at all. The productivity of these workers will be low and probably results in lower profits. At the macro level, mismatches usually lead to lower productivity, lower innovative output, slower adoption to technological developments, and in the end the misuse of human capital (Brunello & Wruuck, 2021).

As explained, due to the increasing demand for a workforce with more education (Handel, 2020), and the ambition to develop into a higher educated population, the concept of a vertical mismatch based on education is a relevant topic. Due to this ambition, the availability and quality of education is increasing. As a result, the incidence of overeducation is also growing, which is shown in Figure 1.2. On top of that, McGuinness, Bergin, and Whelan (2018) indicate that overeducation seems to be an enduring problem and in their results they find it is gradually growing trend. Therefore, this thesis will focus on overeducation.

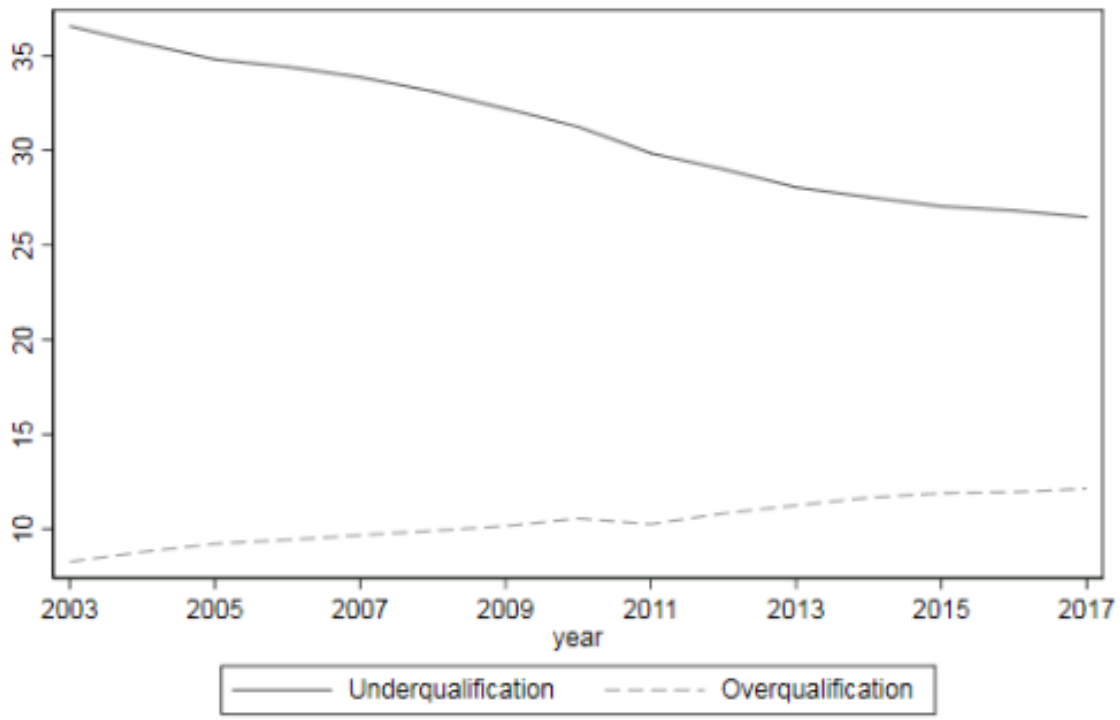

Figure 1.2.: Evolution of underqualification (undereducation) and overqualification (overeducation) in the EU (Vandeplas & Thum-Thysen, 2019)

Studies have shown that mismatch, specifically overeducation, can affect wage (Hartog, 2000; Leuven & Oosterbeek, 2011; McGuinness, 2006). The problem of overeducation affecting wage lies in the consequences of its effect as it can lead to income inequality, economic inefficiency and less returns on education. For example, Brunello and Wruuck (2021) explain how a overeducation is a big factor contributing to the increasing wage dispersion in the United States. Furthermore, wage premiums compared to properly matched workers can cause employers to invest less, eventually leading to less innovations. Also, the mentioned decreasing return on education arises as a problem for individuals and society (Carroll & Tani, 2013; Shahidan & Ismail, 2021). Education, the availability and quality of it, is often subsidized by public funds. When the outcome of this investment, an educated person, is underutilized, the returns on education will decline. In the end, the time and money the student used and the public funds are underutilized. Hence, understanding the relation between overeducation and wage is an important step in order to eventually address these challenges. So, a better understanding of this relation is, besides the academic knowledge, also important for policymakers, employers, and individuals seeking to mitigate the negative consequences.

The Netherlands is an interesting case to study the effect of overeducation on wage. According to van der Mooren and de Vries (2022), the amount of people with a higher education in the Netherlands is growing. This trend is also found by Statistics Netherlands, a Dutch organization that collects and publishes reliable statistical information about the Netherlands. They explain that the proportion of people between the ages of 15 and 75 with a higher education has increased over the past 40 years, from 1 in every 9 people in 1981 to 1 in every 3 people in 2021 (CBS, 2022). During the same period, the amount of people with lower education has declined from 58% to 26%. Both trends combined will make for a perfect place for overeducation to occur. An interview by Lucas (2022) with a sociologist of Statistics Netherlands reveals that these trends can be explained by policies. For example, policies created to reduce the amount of people that leave education early. But also the role of the knowledge economy is mentioned, where policies are made to increase the level of education of the population. These trends create an environment where overeducation can prosper. Therefore, it is not unexpected to find that jobs, previously occupied by lower educated people, are now done by higher educated people (Lucas, 2022). Nonetheless, not many studies have specifically focused on the Netherlands.

Furthermore, a lot of studies have examined overeducation's effect on wage, but research exploring how this phenomenon varies across certain industries and occupations has been limited. In this context, industries refer to groups of entities that carry out the similar types of activities. Occupations are categorized based on the skill level and specialization that is required for the job. Also, the same occupation can exist in different industries. For example, an administrative assistant (occupation) may be found in both the manufacturing and the market services industries. When the differences in wage effects across industries and occupations are unclear, policymakers, employers, and employees might struggle to mitigate the negative consequences. For instance, a certain occupation might experience a larger wage penalty for overeducation compared to properly matched workers with the same education. In such cases, a uniform policy targeting all occupations may not be sufficient to address the problem for the occupation with a larger wage penalty.

1.2. RESEARCH OBJECTIVE

Section 1.1 has explained the challenges due to overeducation and the effect on wages. It also described why the Netherlands is an interesting case for this topic and why the wage effect across industries and occupations should be examined. So, to address this knowledge gap, the research objective of this research is to investigate and analyze the effect of overeducation on the wage of workers across occupations and industries in the Netherlands. By exploring how overeducation affects wages across different industries, this research can provide insights that go beyond the general relationship between overeducation and wages. By doing so, this research will contribute to a comprehensive understanding of the effect of overeducation on wages. Afterwards, the knowledge gained by this research can be developed further and used to aid in mitigating the effects of this type of mismatch.

1.3. RESEARCH QUESTIONS

To investigate the problem and reach the objective described in Sections 1.1 and 1.2, the main research question is established:

Main RQ: *To what extent does overeducation's effect on the wage of a worker vary across occupational groups and industries in the Netherlands?*

To help in answering the main research question and guide the research, the following sub-research questions are formulated and will be answered in chronological order:

SRQ1: *How is overeducation defined and measured?*
This question will look into how overeducation is defined and measured. Also, the consequences when a specific measure is chosen are addressed.

SRQ2: *What variables influence the relationship between overeducation and wage?*
When one wants to investigate the relationship between an independent variable and a dependent variable, it is important to investigate which other variables might also influence the dependent variable or the relationship between hem. These can then be used as control variables to get a undistorted representation of the examined relationship.

SRQ3: *Which industries and occupations are affected the most by overeducation and its effect on wage in the Netherlands?*
This question will examine the differences of the impact of overeducation on wage across different industries and occupations. Understanding how the effect (possibly) differs for these factors will provide insights into whether certain industries or occupation types are more affected by overeducation and its effect on wage than others, guiding policymakers on where interventions might be necessary.

1.4. RELEVANCE OF RESEARCH

This section will discuss the scientific relevance, practical relevance, and the relevance to the study Management of Technology (MOT).

1.4.1. SCIENTIFIC RELEVANCE

This thesis will contribute to the academic fields of wage determination and labor economics. By investigating overeducation's impact on wage, it tests existing theories that explain the effect of overeducation on wage, specifically the human capital theory by Becker (2009), the job competition theory by Thurow (1975), and the assignment theory by Sattinger (1993). Additionally, incorporating industries and occupations as variables offers new insights into if a certain theory still holds across all industries and/or occupations. These differences across industries and occupations and their incorporation in models are not extensively researched yet. Furthermore, the size of the wage effect of overeducation is compared across industries and occupations.

1.4.2. PRACTICAL RELEVANCE

Investigating the effect of overeducation on wage is important for multiple parties, namely for employees, employers, and policymakers. Individuals close to entering the labor market need to understand how overeducation might impact their income and how the effect differs across different industries and occupations. Also, even before selecting a study, they can use the knowledge about the wage effect for a certain occupation and/or industry to make more informed decision about their education path. This allows them to assess the return on investment of their education in terms of both time and money. Employers are often searching for maximum profitability. When overeducated employees require a higher income compared to properly matched employees, the profitability of the firm will decrease. By understanding the impact of overeducation on wage for their industry and the type of occupation and productivity and compare them, the employer can make a better informed decision and reach a better return on their investment (Vandeplas & Thum-Thysen, 2019). Furthermore, education is subsidized by the government as it is their goal to advance into a more highly educated population. When overeducation occurs more often, knowledge of the effect of overeducation on wage can help in altering their policies to mitigate the negative consequences. For example, policymakers can change their funding decision, possibly resulting in higher returns on public resource investments. By understanding the variations of the effect of overeducation on wage across occupations and industries, policymakers will have more insights where policy changes are necessary.

1.4.3. RELEVANCE TO MANAGEMENT OF TECHNOLOGY

This thesis is done in partial fulfilment of the requirements for the degree of Master of Science in Management of Technology (MOT). During the Management of Technology (MOT) program at TU Delft, one will learn to examine and understand technology as a corporate resource. This is done through, among others, addressing economics, finance, and organizational strategy, as these factors play a role in enhancing the technological and innovative capabilities of firms.

This study researches the effect of overeducation on wage across certain industries and occupations. Understanding how overeducation impacts wages and how this effect varies will help firms to manage human resources more effectively by reducing inefficiencies related to overeducation. By efficiently allocating their resources because of the generated knowledge of this thesis, firms will have more corporate resources available for innovation and technological development. Using the same knowledge, policymakers can make better decision to prevent overeducation, for example by changing policies. Both will help to improve corporate efficiency, leaving more resources available for technological developments and innovations. Therefore, this thesis intersects with technology and is relevant for the MOT program. In order to conduct this thesis, knowledge and analytical and research skills obtained during the MOT curriculum is used.

1.5. THESIS OUTLINE

In order to get a better overview of this thesis, the structure of it will be explained. In the next section (Section 2) a literature review is conducted. It will contain the current literature about overeducation and how to measure it. Next, the current literature on overeducation and its effect on wage are examined, including other variables that influence this relationship. The literature review will end with a conclusion containing why this research is relevant and how it adds to the existing literature. So, Section 2 will give answers to sub-research question 1 and 2. In Section 3, the gathered knowledge from the literature is used as evidence to create a theoretical framework for this thesis, after which the hypotheses will be developed. These hypotheses will eventually be tested in this thesis. Before testing, the research strategies will be defined in Section 4. In the following section, Section 5, information is given about which data will be used and how it will be analyzed. Section 6 reports the results of this thesis. This contains the outcomes of the descriptive statistics and the regression models, which evaluate the earlier developed model. The final results and insights will be discussed and clarified in Section 7. This section will also contain the theoretical contributions, practical implications, limitations of the research, and future research recommendations. Lastly, Section 8 will conclude this thesis by laying out the main conclusions of this thesis using the answers of sub-research question 1-3 to answer the main research question.

2

LITERATURE REVIEW

This literature review will explore the existing literature about overeducation and wage. Then, in Section 3 the gathered evidence will be used to base the theoretical framework on. The methodology that is used to find relevant sources is explained in Appendix A.
This literature review aims to answer the following sub-research questions:

SRQ1: *How is overeducation defined and measured?*

SRQ2: *What variables influence the relationship between overeducation and wage?*

First, the different types of mismatches are reviewed. Then, literature will be examined about how one can measure overeducation and the (dis)advantages of these measures. Next, the theories that explain the effect of overeducation on wage are discussed. Then, models that are used to determine the wage based on education are described. Furthermore, existing empirical evidence on the relation between wage and overeducation is discussed.

2.1. TYPES OF MISMATCHES

When one wants to investigate mismatches and related concepts, it is important to get a clear understanding of what the different types of mismatch are and how they differ from each other. This necessity is illustrated by the disagreements among studies about the concept of skill (OECD, 2017). A mismatch can happen based on the skills of a worker and those required for a job. Different studies use different ways of describing skill. This makes it difficult to understand the concept of skills and reach a general consensus about how it is defined and used. Hence, a good understanding of the concept of mismatches and the distinction between educational mismatch and the other types of mismatches is essential. Only then it is possible to contribute to the existing literature and theories (Shahidan & Ismail, 2021).

Figure 2.1 shows how mismatch can be divided into an imbalance between demand and supply of labor and a job-education mismatch (i.e. occupational mismatch). The mismatch between demand and supply of labor is determined through the difference between the demand and supply based on comparing the working and unemployed people. However, this doesn't necessarily represent the actual mismatch of a specific person. An example of this imbalance would be that there are more vacancies for certain jobs than job-searching workers. The job-education mismatch refers to a mismatch between the education a specific worker has acquired and the education that is required by a specific job. This can be broken down further into a horizontal or vertical mismatch. The term horizontal relates to a difference in the field of study of the acquired education of a worker and the field of study required by a job. For example, a person might have studied in the field of engineering, but ends up in a financial analyst function. The vertical mismatch explains when there is a mismatch between the level of study of the person and the level of study required by the function. This vertical mismatch can be divided further based on skills and education. An example of a mismatch based on skills is when an employee has skills that are not required by the job. In this case, the worker is classified as overskilled. On the contrary, when a person does not have certain skills that are required to properly function, that person is underskilled. This has demonstrated how these types of mismatches differ from a mismatch based on education and what is not encompassed by the term educational mismatch. The focus of this thesis is on overeducation, as explained in Section 1. Therefore, the remainder of this section will only cover overeducation (and its wage effect).

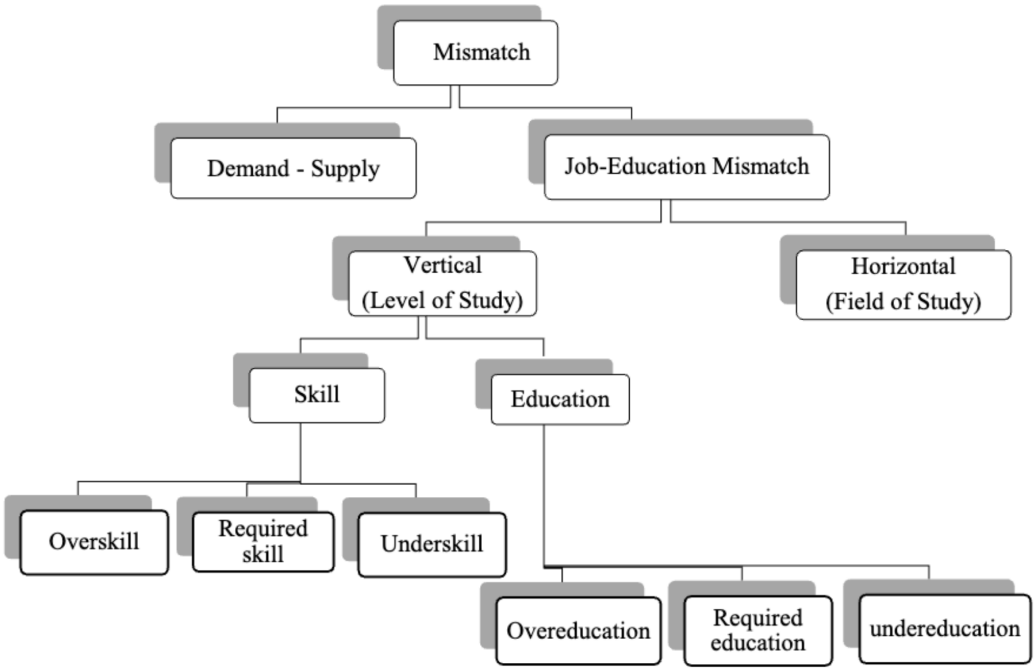

Figure 2.1.: An Overview of the Mismatch Types (Shahidan & Ismail, 2021)

2.2. OVEREDUCATION

The vertical mismatch literature has mostly been dominated by research into overeducation (McGuinness, Pouliakas, & Redmond, 2018). Yet, there are still gaps to be filled in the existing literature. For clarity the definition of the mismatch based on education is explained. This type of mismatch compares the education of a worker with the education required by a job. When properly matched, the worker has obtained only the required education. Furthermore, overeducated workers have more education than their jobs require, while undereducation refers to the opposite situation. As explained in the introduction (Section 1), this type of mismatch is impacted by the ambition to obtain higher education and changing landscape of labor (Handel, 2020; van der Mooren & de Vries, 2022), which can make this very complex.

2.3. MEASURING OVEREDUCATION

The definition of overeducation is clear and concise. However, different approaches are used to measure it. Studies have commonly discussed three ways of measuring overeducation: self-reported, realized-matches, and the job evaluation method, also known as the subjective, empirical or statistical, and objective approach respectively (Brunello & Wruuck, 2021; European Training Foundation, 2019; McGuinness, Pouliakas, & Redmond, 2018; Shahidan & Ismail, 2021). These measures each have their pros and cons. In the end, Shahidan and Ismail (2021) explain that the all of the methods of measurement can be used, keeping in mind their limitations, and the choice is dependent on the availability of data.

Self-reported. The 'self-reported' method is based on the assessment of the mismatch by workers themselves. Hence it is a subjective measure based on the workers' perceptions. An advantage of this approach is the ease of use in surveys, but bias is easily incorporated in the answers of employees causing measurement errors. The source of these measurement errors, according to Brunello and Wruuck (2021), is that persons do not have a accurate view of job requirements as they often exaggerate them (McGuinness, 2006). As a result, the requirements of their jobs are higher, and the probability of being overeducated is lower. Then, there is a measurement error between the real requirements of a job and the requirements as described by the worker. As a result, McGuinness, Pouliakas, and Redmond (2018) describe that the incidence of overeducation is lower when the subjective measure is used due to these measurement errors. This is not unexpected. Another issue

related to this approach is the lack of uniformity in terms of the overeducation questions asked within datasets. Therefore, it is difficult to compare the results between them (Leuven & Oosterbeek, 2011). How a question is being asked might invoke a certain answer. When these questions differ, a variety of answers is found.

Realized-matches. The 'realized-matches' method compares the requirement of a job, based on the average or mode of the level of education of the employees that perform this job, to the education level of an employee. The incidence of overeducation can often easily be computed using existing data with this approach. A drawback is that it uses the mode or average education level of the employees, which both distort the measurement. The average or mode of the educational attainment does not have to represent the actual required education level for a specific job. The realized-matches approach assumes that the distribution of the current educational attainment of the workers of a specific job accurately represents the requirements of the job. This assumption might not always be true. To illustrate, Lucas (2022) explains that lower educated people are now replaced by higher educated people for the same job. The requirements of the job don't change, but the distribution does due to the increase in higher educated workers.

Job Evaluation. The third method uses experts to assess the educational requirements of a specific job, which increases objectivity compared to the self-evaluation method, but it keeps some level of subjectivity and it also is expensive. For this approach all jobs are classified. However, McGuinness, Pouliakas, and Redmond (2018) explain that there are difficulties because of the different interpretations and expectations of employers for each classification. Therefore, this approach usually uses the classifications determined by the Dictionary of Occupational Titles in the International Standard Classification of Occupation (ISCO) (Shahidan & Ismail, 2021). Quintini (2011) describes that the use of this classification system assumes that all jobs with the same titles require the same level of educational attainment, which is true when countries use the same classification system for occupations. However, this assumption can also be false. There might be small differences for the requirements of the same occupation, but done in different companies or industries. In this case, the assumption does not hold and the method is also not reliable.

The literature on these methods explains the limitations and strengths of each of them. The self-reported measure offers ease of use but suffers from biases and lack of uniformity. The realized-matches approach uses data to determine the requirements of a job, but the average or the distribution might not represent the actual requirements. Experts are used for the job evaluation method, which results in a structure due to the classifications. However, subjectivity remains a problem because of the human aspect. Given these pros and cons, the decision for a measurement method must be made, mainly based on the data that is available. This study will use one of these measures. The choice is explained in Section 5. Future research on this topic should also focus on finding a consensus on how the requirements are determined, objectively and uniformly, to be able to investigate the impact of overeducation on wage without bias and measurement errors. Now that the measures of overeducation are discussed, the focus shifts towards the effect of overeducation on a worker's wage.

2.4. OVEREDUCATION'S EFFECT ON WAGE

A lot of studies have examined the effect overeducation has on the wage of a worker. There are theories that try to explain how and if overeducation has an effect on wage. These theories are tested with models which determine the wage of a worker based on education. Both the theories and models are discussed in this section.

2.4.1. THEORIES EXPLAINING OVEREDUCATION'S EFFECT ON WAGE

Some theoretical models will give us insights into the incidence of overeducation and how it influences wages. Three models are commonly mentioned in the current literature: the human capital theory (Becker, 2009), the job competition theory (Thurow, 1975), and the assignment theory (Sattinger, 1993).

Human capital theory. The human capital model examines influences of the supply side of the labour market. Becker (2009) proposes that workers will receive a wage that reflects the amount of output

they deliver. This output is dependent on the human capital of a worker. Human capital consists of the knowledge, skills, and competences of a person. In this case, the phenomenon of overeducation and its wage effect presents a challenge to this theory. Overeducation occurs when an individual possesses more education than their job requires and it is found that overeducated workers often face a wage penalty. Contrary to what is expected by the human capital theory, this suggests that the additional education, which is additional human capital, does not always translate into a higher wage. The observed wage penalty for overeducation suggests that the relationship might be difficult to interpret. The explanation is that the wage penalty may be caused by factors that are not captured or cannot be captured in some wage calculations, such as unobserved human capital. For example, overeducated individuals might have lower levels of other skills or abilities that are not directly related to education but are gained through work experience. These unmeasured skills or abilities may be just as important in determining productivity and wages. The additional years of education will translate into a higher wage, but there are also penalties for lower levels of other abilities. This also highlights the need for enough controls for both these aspects of the supply side, formal education and experience.

Job competition theory. Another model looking into overeducation and how it affects wage is the job competition model. This theory primarily uses the demand side to describe overeducation and its effect on wage. According to Thurow (1975), only job characteristics are used to determine the wage (and productivity). Job-searching individuals are looking for jobs that come with higher wages. Furthermore, employers want to aim for a high productivity, which can be obtained through workers. The workers and jobs are ranked by educational attainment and wage respectively. In the end, the people with the highest level of education will get the job with the highest wages. Because of the increasing number of higher educated people, more people are overeducated for their jobs. And in the case of the job competition theory, the wages are determined by the characteristics of the job. So, the wage penalty occurs as a result of wages being determined by the job requirements.

Assignment theory. The last theory is called the assignment theory. This theory focuses on the matching process between workers and jobs, in which aspects of both the demand and supply side are considered (Sattinger, 1993). Workers and jobs are matched in order to maximize productivity. When they are properly matched, it will positively impact productivity and wages. On the other hand, bad matching results in a negative effect on productivity and wage. Thus, in this case, the theory describes how both the job requirements and level of education of the worker influence his/her wage. Then, overeducation and undereducation indicate a bad match, resulting in lower returns on education.

In summary, three theories have been described that provide us insights on the different viewpoints on how overeducation has an impact on wage. The human capital theory emphasizes on the supply side, the competition model predominantly uses the demand side, and the assignment theory combines both supply and demand to look at the quality of the match. At this moment, there has not been a general consensus on what theory best describes the the actual relationship between overeducation and wage, but the effect on wage is investigated by multiple studies. This study will examine which theory is true by using the data to determine which of these theories reflects the actual situation in the Netherlands best. The next section will also explain how the returns on (over/under-)education are expected to be when a certain theory is true. Then, after our analysis, the calculated returns can be examined and eventually used to determine which theory best describes the data under consideration.

Furthermore, insights of the theories are used to create the conceptual model. Both the supply and demand side are used in this study. For example, in line with the human capital theory, the experience of a worker will be used as a control variable. The required years of education for a job are used, looking at the demand side.

2.4.2. MODELS FOR DETERMINING WAGE BASED ON EDUCATION

Different studies have researched the relation between overeducation and wage to determine which theory, mentioned in Section 2.4.1, is true. To be able to determine which theory best represent reality, models are used that calculate the wage of a worker based on the education and other variables. Thus, models are used as a tool to test and improve the theories and hypotheses. It does not provide an explanation of the wage effect of overeducation on its own. In studies two models are commonly

used, the ORU model and the Verdugo & Verdugo model. Often, causal effects were found with these equations. Still, these causal effects are said to be invalid due to the omitted variable bias and measurement errors, according to Leuven and Oosterbeek (2011). First, the two models are briefly explained.

ORU model. The wage model based on Duncan and Hoffman (1981), which is also called the 'workhorse model in overeducation', estimates the effect of overeducation (O), required education (R), and undereducation (U) on the wage. This equation is widely known as the ORU model. Hartog (2000) uses this model and describes the following econometric model (Equation 2.1):

$$lnw_i = \beta \cdot S_i^o + \beta \cdot S_i^r + \beta \cdot S_i^u + \beta \cdot x_i + \epsilon_i \tag{2.1}$$

where w_i represent the wage of an employee, S_i^r are the years of education required, S_i^o are the number of years of overeducation, S_i^u are the number of years of undereducation, x_i represents other independent variables, and ϵ_i is a random error term. The $\beta's$ are the returns on the corresponding variables. Equation 2.1 is a variation of a traditional wage equation, the Mincer wage equation. The ORU model differs from the traditional Mincer wage equation as the requirements for a job are also used, which is not the case in the traditional Mincer wage equation. The Mincer wage equation only uses the amount of obtained education (Mincer, 1974).

Verdugo & Verdugo model. Another model that is often used in this field of labor economics is a model described by Verdugo and Verdugo (1989). Instead of using required education of a job as control variable, they use the years of education completed. Equation 2.2 represents this model (Verdugo & Verdugo, 1989).

$$lnw_i = \beta \cdot EducationYears_i + \beta \cdot OE_i^{Dummy} + \beta \cdot UE_i^{Dummy} + \beta \cdot x_i + \epsilon_i \tag{2.2}$$

In this model the wage of an employee is indicated by w_i, completed years of education by $EducationYears_i$, OE_i^{Dummy} is a dummy variable indicating if the worker is overeducated, UE_i^{Dummy} is a dummy variable indicating if the worker is undereducated, x_i represents other control variables, and ϵ_i is a random error term. Again, the $\beta's$ are the returns on the corresponding variables. Thus, the model by Duncan and Hoffman (1981) compared workers who have the same job (requirements) and Verdugo and Verdugo (1989) compared workers with the same educational qualifications. The former model finds a wage premium for overeducated workers compared to workers with the same job and the latter finds a wage penalty for overeducated workers compared to workers with the same education. So, although it might seem like the two findings are opposites, this is not the case as they use a different reference point. A critique of the model by Verdugo and Verdugo (1989) is the fact that they use dummy variables for over- and undereducation (Leuven & Oosterbeek, 2011). The amount of years of overeducation or undereducation is not taken into account. Hartog (2000) mentions that the deletion of this equation would result into a body of research that is easily comparable. Hence, because of the critiques, this research will examine the wage effect of overeducation using the ORU model.

Let's take a closer look at the ORU model as it is important to understand the model and know how to interpret the results. For clarity, the ORU model is described by Equation 2.3, which is restated with more detail than Equation 2.1.

$$lnw_i = \beta_1 \cdot S_i^o + \beta_2 \cdot S_i^r + \beta_3 \cdot S_i^u + \beta \cdot x_i + \epsilon_i \tag{2.3}$$

The literature indicates that this equation is often used to determine the wage effects of overeducation. This study will also use this model in order to determine the wage effects and check which theory is true. Therefore, the next questions that arise are: how one can interpret the ORU model and how can one check which theory is true from the beta coefficients and variables in the equation? To start, the ORU model uses required years of education as a variable, making it look like only the workers with the same required education are compared. Still, one can also make a conclusion about the wage effect compared to a worker with the same education. To interpret the the wage effect using the ORU model, β_1, β_2, and β_3 are examined. β_1 represents the return on a year of overeducation. If this coefficient is positive, an additional year of overeducation does result in a higher wage. β_2 is the coefficient belonging to the required years of education. This indicates how much a year required education changes the wage. The coefficient belonging to undereducation, β_3, is usually negative. It

is expected that both β_1 and β_2 are positive, as the additional education will probably result in some increase in wage. It is important to note that it is expected that β_1 is smaller than β_2. Otherwise, the surplus years of education will increase the wage more or equally than the initial required years of education by the job. Whether this is actually the case or if overeducated workers do not experience a wage premium compared to workers with the same job has to be determined.

Although it might seem like one cannot indicate whether there is a wage penalty for overeducation compared to properly matched workers with the same education level, this is not the case. This penalty is presented by the different sizes of the coefficients that are often found in the studies examining the wage effects of overeducation. The coefficient for years of overeducation, β_1, is often smaller than the coefficient corresponding to required years of overeducation, β_2. Then, years of education that are not properly used, years of overeducation, do not have the same return as years of education that are properly used, required years of education. The returns are lower for a year of overeducation. Using the difference between the coefficients, workers with the same education can be compared. A person who is overeducated divides the obtained years of education over the variables belonging to required years of education and years of overeducation. On the other hand, a person who is properly matched with the same obtained years of education converts all those years into required years of education. Because the coefficient of overeducation, β_1, is smaller than the coefficient for required years of education, β_2, the properly matched worker will have a higher wage when all other factors are the same. This illustrates how one can determine if there is a wage penalty for overeducated workers compared to people with the same education.

The three theories that are explained in Section 2.4.1 will be tested using the ORU model. To determine whether a theory is true, again the coefficients for overeducation, undereducation, and required years of education are examined. The human capital theory describes that the wage is determined by the human capital of a worker. In the case of education, more years of obtained education will increase someones wage. There should be no difference between the returns on overeducation and required years of education. As together these variables are the obtained years of education of a worker. More education will only increase the wage. There is no wage penalty for years of overeducation. Hence, according to the human capital theory, the coefficients β_1 and β_2 should be the same, and the coefficient for years of undereducation is in the opposite, negative.
The job competition model explains how wage is determined by the demand side. The wage is completely determined by the requirements of the job. Then, the coefficients for over- and undereducation, β_1 and β_3 respectively, are both equal to 0. Only the coefficient belonging to required years of education, β_2, is positive.
The assignment model looks at the quality of the match. A bad quality of the match based on education results in a lower wage than when a worker is properly matched. In this situation, none of the coefficients for overeducation, undereducation, and required years of education are equal to each other and are also not equal to zero. β_1, should be positive but smaller than β_2. Overeducation indicates a bad match but it does provide some increase in wage. β_3 should be negative as it also indicates a bad quality of the match.
After the analysis of this thesis is performed and the results are described, a conclusion can be drawn which theory best describes the effect of overeducation on wage and whether there are deviations from a theory for certain industries and/or occupations.

2.4.3. CONTROL VARIABLES

Leuven and Oosterbeek (2011) determines that there are still some problems in determining the wage effects of educational mismatch. These problems are related to the omitted variable bias and measurement errors. The omitted variable bias is a major problem related to the overeducation literature, as it is in many social science studies. The estimates often suffer from omitted variable bias (Leuven & Oosterbeek, 2011). Omitted variable bias refers to variables, which are not included in a model but do explain the dependent variable and, at the same time, are correlated with the independent variable(s). Sometimes these variables are known, but cannot be included in the model for several reasons (e.g., no data available). An example of an omitted variable is intrinsic ability. Someone could have lower intrinsic abilities, which is correlated with both overeducation and wage. Having better abilities will probably result in a higher level of education and thus overeducation, but also a higher wage. Hence, when intrinsic ability is not included, the effect of overeducation on wages also contains an effect which is created by intrinsic ability. Resulting in an upward bias,

as part of the wage effect due overeducation is caused by the intrinsic ability of a person. Hence, part of the additional wage is not the consequence of overeducation and therefore the wage effect is overestimated. The problem of omitted variables bias indicates the importance of control variables for inference. However, to much control variables can also distort the relationship. Different empirical studies have examined and used different control variables. These studies will be described in the next section to get an overview of which variables are used and should be used in this research. When the conceptual model is created in Section 3, the variables are included as control variables.

2.5. REVIEW OF EMPIRICAL EVIDENCE

It is now known how one can measure overeducation, why and how overeducation has an effect on wage according to theories, and which model can be used in order to test the theories. This section will review the empirical evidence on the topic of overeducation and wage to get a better understanding of what is researched and what is not yet researched. A study by Choi *et al.* (2020) has listed the focus of recent cross-country studies which use data from a survey by the Programme for the International Assessment of Adult Competencies (PIAAC) to analyze educational mismatch. Figure 2.2 shows this list.

Author(s)	Type of mismatch	Focus
Desjardins and Rubenson (2011)	OE, UE, OS, US	Causes and differences between OE and OS.
Allen, van der Velden, and Levels (2013)	OE, OS	New measures. Incidence of OE and OS.
Pellizzari and Fichen (2013)	OE, OS	New measures. Incidence of OE and OS.
Flisi et al. (2014)	OE, OS	Predicted probabilities for occupational mismatch types.
Levels, van der Velden, and Allen (2014)	OE, UE, OS, US	Relation between wage differentials and mismatches.
Maer Matei (2014)	OE, UE, OS, US	Incidence of OE and OS.
Perry, Wiederhold, and Ackermann-Piek (2014)	OS, US	Implications of skill mismatch on labour markets.
Quintini (2014)	OS, US	New measures. Incidence of OS and US.
Adalet McGowan and Andrews (2015)	OS, US	Effects on public policies.
Maršíková and Urbánek (2015)	OE, UE	Incidence of OE and UE.
Pelizzari et al. (2015)	OS, US	TRE skill mismatch and ICT knowledge.
Pouliakas and Russo (2015)	OE, OS	Link between job tasks and cognitive skills.
Vera-Toscano and Meroni (2016)	OE, OS	Evolution of occupational mismatch in Europe.
Flisi et al. (2017)	OE, OS	New measures. Incidence of OE and OS.
Adalet McGowan and Andrews (2017)	OE, UE, OS, US	Effects on productivity.
Chłoń-Domińczak and Żurawski (2017)	OE, OS	New measures. Effect of mismatch at sectoral level.
Cim, Kind, and Kleibrink (2017)	OE, OS	Mismatch disparities between immigrants and natives.
Pellizzari and Fichen (2017)	OE, OS	New measures. Incidence of OE and OS.
Mateos-Romero, Murillo-Huertas and Salinas-Jimenez (2018)	OE, OS	Relation between wages and mismatches.
Mateos-Romero & Salinas-Jimenez (2017)	OE, OS	Effects on job satisfaction.
Montt (2017)	OE, UE, OS, US	Occupational mismatch and field of study mismatch.

OE, overeducation; UE, undereducation; OS, overskilling; US underskilling.

Figure 2.2.: Overview of cross-country studies using PIAAC data (Choi *et al.*, 2020)

The figure illustrates that PIAAC data has not been used to examine the effect of overeducation on wage, with the specific focus on the influence that industries and occupational groups have on this relationship. At first, it might look like Chłoń-Domińczak and Żurawski (2017) cover the effects of sectors and occupational groups on overeducation's effect on wage. However, they only look into the incidence of overeducation across industries and occupations. The only studies that specifically focus on the effect on wage using PIAAC data are studies by Levels *et al.* (2014) and Mateos Romero *et al.* (2017). The former study focuses on how the effect of educational mismatch on wage can be understood. This study used the ORU model in order to determine the extent to which skills influence the impact of overeducation on earnings. During this study they also looked at the returns on

education. Let us also go through the studies that did not use PIAAC data, starting at the foundation of the ORU model.

One of the first studies that became quite influential is the study by Duncan and Hoffman (1981). They used the ORU model to investigate the wage effects of overeducation using data from the US work force. It found that overeducation does have economic value for the worker, as it results in a wage premium compared to workers with the same job. However, more value can be gathered by being properly matched (Duncan & Hoffman, 1981). They also looked at the incidence of overeducation across job types, but did not examine the impact of the occupation type on the wage effect of educational mismatch.

Hartog (2000) progressed with the works by Duncan and Hoffman (1981) and used their model to map the incidence of overeducation for multiple countries, and its consequences for earnings. By using the ORU model they established the returns to overeducation on be half to two-thirds of the returns on the required education (Hartog, 2000). This time occupational dummies are used to look include its effect on wage, but not its effect on overeducation's effect on wage.

A study that took a slightly different approach is a study by Allen and van der Velden (2001). They wanted to show the effect of educational mismatch and skill mismatch on wages and job satisfaction. It is concluded that the educational mismatch is a stronger determinant of wages compared to skill mismatch. Only a small amount of the wage effect is caused by skill mismatch (Allen & van der Velden, 2001).

So far, these studies have examined the wage effects of overeducation and looked at differences between firms and found similar results. Groeneveld and Hartog (2004) want to add to the field of overeducation to investigate whether the same conclusion can be drawn using data from a single firm. They also use the ORU model, but are one of the first to incorporate the hiring standards of a firm. According to Groeneveld and Hartog (2004), this way of measuring improves the fit of the model. Furthermore, similar results are found as before.

Following Rubb (2006), overeducated individuals earn less than similarly educated individuals who are properly matched in their jobs but more than workers with similar jobs who are properly matched. Again, these findings are found in several studies before. Moreover, this study examined the effect of experience on the wages of workers. They found that experience positively influences wages in the beginning, but as one gets older the returns on experience diminish (Rubb, 2006).

An important article in the educational mismatch literature covers research by Verhaest and Omey (2006). The previously mentioned studies have used different measures of overeducation. This paper examines the differences between these measures regarding the effect of overeducation. It is determined that the different measures cause the magnitude and significance of the effect of overeducation to diverge quite a bit. This has to be taken into account in research on the topic of educational mismatch. They also included industry and professions as control variables. Yet, they did not focus on the effect of these variables on the size of the effect of overeducation on wage (Verhaest & Omey, 2006).

Besides cross-sectional studies, some studies have also included the effects over time. An example of such a study is a study by Korpi and Tåhlin (2009). They wanted to test their hypothesis regarding the diminishing effects of educational mismatch as time goes by, meaning that the impact on a worker's wage will approximate zero over a employee's career. However, this hypothesis is not supported by the data. Again the same effect regarding overeducation on wage is found as in the studies mentioned before.

A lot of the studies have used the ORU model to examine the effect of overeducation on wages. Leuven and Oosterbeek (2011) explain that it is extremely difficult to establish accurate causal effects of being overeducated or undereducated due to the omitted variable bias, even though a lot of these studies have found similar results. The measurement error of measuring the amount of overeducation also adds to the complexity of this problem (Leuven & Oosterbeek, 2011).

Iriondo and Pérez-Amaral (2013) aim to tackle the problem of omitted variable bias and measurement error. They did this by using consistent estimators. Their results confirmed that overeducated workers

suffer a wage penalty compared to properly matched workers. An important addition of this research is that they found that in the early stages of a worker's career, the wage is determined by the level of education obtained. Later in the careers, educational requirements explain the differences in wage better.

The study by Mateos-Romero and Salinas-Jiménez (2018) analyzes the effects of labor mismatches on wages and on job satisfaction using PIAAC data of several countries. Once again the ORU model is used and the results indicate that educational mismatch causes greater effects on wages. The findings also highlight that the ability or skills have a very minor effect on wage. The effect across all industries and occupational groups is not examined.

Skills and their effect on wage are investigated more often in this time period. Nieto and Ramos (2017) examine the skill heterogeneity theory, which could partially explain the differences between wages of overeducated workers and properly matched workers. This is determined to be the case.

Sellami *et al.* (2017) look at different factors that represent the unobserved heterogeneity. Their results show that overeducated workers without a field of study mismatch earn less than adequately educated workers with the same obtained education. The individuals who are working outside their field of study do not always experience such a penalty.

The last study that is discussed is a study by De Santis *et al.* (2022). They find a wage penalty for severely overeducated, the penalty is bigger than those who experience a mild level of overeducation. In general the overeducation wage penalty is confirmed, but there is great heterogeneity among overeducated observations in this study. Respondents that have relatively high wages experience a much lower penalty, or even a premium in some cases. Finally, the examined effects of a horizontal mismatch based on knowledge is not statistically significant.

This part of the literature review has shown that the ORU model is often used to examine the effect of overeducation on wage. However, the research in the last 10 to 15 years has focused on the unobserved heterogeneity of overeducated workers, which has been determined to be a problem in the overeducation literature. A wide variety of variables have been examined. Yet, no study focuses on the impact of occupational groups or industries on the wage effect of overeducation. Lastly, it is important to note that these sources are just a subset of the overeducation literature, but it gives an overview of what is already researched on this topic.

Leuven and Oosterbeek (2011) have summarized the results in studies examining the returns on overeducation, undereducation, and required years of education. These studies used the ORU model to estimate the effect. It is found that the average return on the required years of education (β_2 in Equation 2.3) is equal to 0.089. One additional year of education that is required is compensated with a 8.9% higher wage, keeping all other factors the same. The return on a year of overeducation is 0.043 (β_1 in Equation 2.3) and the return on a year of undereducation is -0.036 (β_3 in Equation 2.3). One more year of overeducation is only compensated with a 4.3% higher wage. The values of the coefficients support the assignment theory, as explained in Section 2.4.2.

2.6. CONCLUSION OF LITERATURE REVIEW

The different ways of measuring overeducation are discussed. Literature explains that the choice for a measure is mainly dependent on the data that is available. Afterwards, three main theories are described which provide an explanation about why and how overeducation impacts wage. The human capital theory suggests that the wage of a worker solely reflects the human capital that workers possesses. Although the human capital is an important determinant of the wage, the characteristics of the job have also been influential for the wage, which is more in line with the job competition theory. Yet, the assignment theory is expected to represent reality the most. Studies have found that both the characteristics of the job and the worker explain the wage. Therefore, variables representing both sides of the labor market will be included in this study to look at the quality of the match. The assignment theory is used as the theoretical framework. Still, the theory needs to be tested using the data under consideration in this thesis to determine if the theory is true. To do so, the ORU model has been found in literature, which will be used as a tool to test the theories and hypotheses.

According to the assignment theory, overeducation has an impact on earnings. Overeducated workers experience a wage premium compared to properly matched workers with similar jobs. However, they could improve their returns on education by being properly matched themselves. These findings will be used as a basis for the hypotheses.

Although studies have looked at the incidence of overeducation across occupational groups and industries, to my knowledge the effect of these factors on the wage effect of overeducation has not been reviewed to be part of the omitted variable bias. It might be the case that the assignment theory does not hold or the wage effect is weaker for some occupation and/or industry types. Also, the review shows that not many studies have specifically focused on the Netherlands (using PIAAC data). Therefore, this study aims to fill these knowledge gaps by examining the effect of overeducation on wage in the Netherlands using the assignment theory and the ORU model. The examination of the impact of occupational groups and industries on this relationship will provide new insights for the wage determination theories based on education.

3

THEORETICAL FRAMEWORK AND HYPOTHESES DEVELOPMENT

This section will briefly discuss the theoretical framework that will be used in this thesis, including the conceptual model. Also, the hypotheses are formulated.

3.1. THEORETICAL FRAMEWORK

As discussed in Section 2.6, this thesis is based on the assignment theory which explains that the wage effect of overeducation is caused by the quality of the match. This theory will eventually be tested by using a variation of the model by Duncan and Hoffman (1981), the ORU model. To fill the knowledge gap that has been found, the aim of this thesis is to find differences in the wage effect across industries and occupations. The standard ORU model includes years of required education, years of overeducation, years of undereducation, gender, and work experience. This research will also include industry and occupation types. Now, the following questions arise which will be addressed next. Why should the wage effect differ across occupation and industry types? Why is the quality of the match more important for certain types?

Occupation types are, in this study, defined as groups of occupations which have the same level of skills and specialization of skills. Overeducation in occupations with a higher level of skills and more specialization could be paid more, as these groups of occupations can leverage their overeducation into a higher productivity (Yeo & Maani, 2017). Furthermore, for some occupations overeducation might not be beneficial as the necessary and value-adding knowledge and skills are not created through formal education, but through experience of working in the occupation.

Industries are delineated according to the type of activities they perform. Some activities might be related to a market, where others to non-market services. Non-market services are often conducted by public organizations. These organizations are regulated. The wages are set in advance, and hence, overeducation will probably get penalized even more. Another factor explaining why the industry type influences the wage effect of overeducation is the R&D intensity of the specific industry. R&D intense industries require a lot of knowledge to be able to innovate and progress. Therefore, these industries will not penalize overeducation as the additional education can be used (in the future) for technological advancements and innovation.

3.1.1. CONCEPTUAL MODEL

The conceptual model that is based on the theoretical framework is presented in Figure 3.1. It contains the control variables, independent variables, and the dependent variable. The arrows represent (expected) causal relations. This research is mostly interested in the effect of overeducation on wage, specifically how it is influenced by occupational groups or industry types. The relationships which are part of the focus of this study are indicated by the red arrows. These arrows are also labelled with the hypothesis they represent. Additionally, the conceptual model is converted to Equation 3.1, which will also be used as the model for the regression analysis. This equation is a variation of the ORU model. Variables that are often found to be correlated with or have an impact on the wage of a worker are included as control variables.

$$ln(Wage) = \beta_0 + \beta_1 \cdot ReqYrs + \beta_2 \cdot OEDU\_YRS + \beta_3 \cdot UEDU\_YRS + \beta_4 \cdot Gender + \beta_5 \cdot Exp + \beta_6 \cdot Exp^2 + \beta_7 \cdot OEDU\_YRS \cdot Ind + \beta_8 \cdot OEDU\_YRS \cdot Occ \quad (3.1)$$

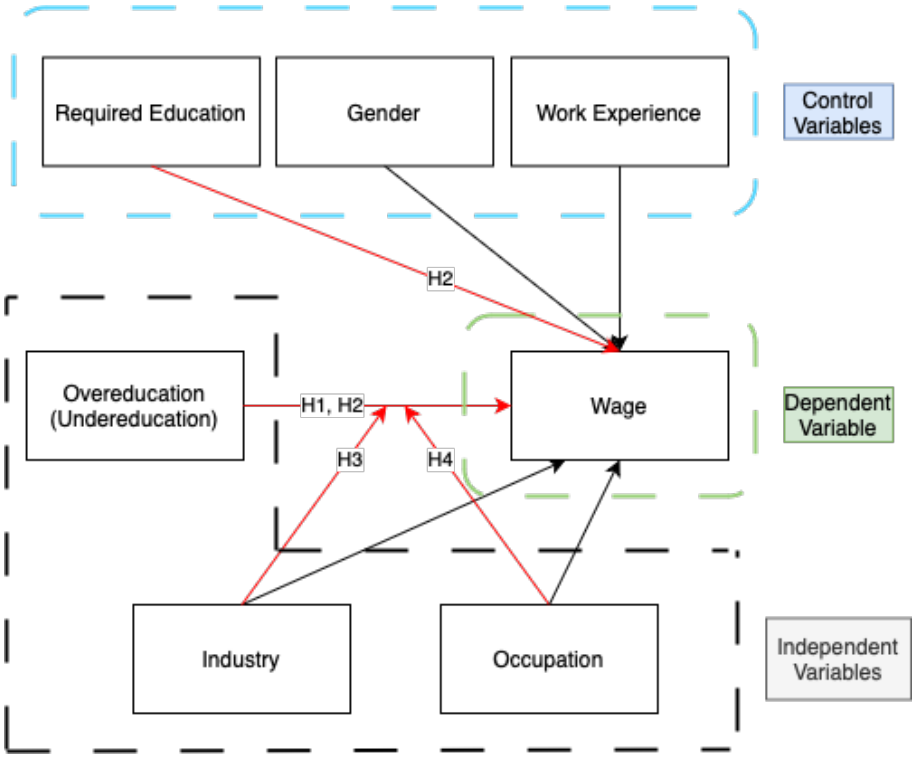

Figure 3.1.: Conceptual model

DEPENDENT VARIABLE

Wage. Wages can be defined in different ways. For example, it can in- or exclude bonuses, hourly or monthly earnings. Mincer (1974) uses the log hourly wages and this is used as a standard in Mincer equations, which our regression models are also based on. Many researchers examining the effects of overeducation on wage use the natural log hourly wages. Hence, this study will also use the natural logarithmic function of wage for several practical reasons.

INDEPENDENT VARIABLE

Overeducation (and Undereducation). Overeducation will have an effect on wage. This relationship is because of the fact that the overeducated worker, compared to a properly matched worker with a similar job, brings more education to the job. This surplus in education is translated into a higher productivity, resulting in a higher wage. Sometimes the additional education can be translated to more or less productivity (and wage) than in other cases. This is indicated by the industry and occupation variables. Certain occupations or industry might influence the relationship between overeducation and the wage of a worker.

Industry Industries can be classified according to the International International Standard Industrial Classifications of All Economic Activities (ISIC), revision 4 (United Nations Statistics Division, 2008). ISIC groups industries together based on the activity they perform. It is expected that some industries might influence the effect of overeducation on wage. The interaction term is

Occupation The International Standard Classification of Occupations (ISCO 2008) groups occupations based on similar skill levels and specializations (International Labour Organization, n.d.). Certain occupations will probably cause a different wage effect of overeducation.

CONTROL VARIABLES

Required years of education. The required years of education is presented as a control variable as the focus of this study is on the effect of overeducation on wage. However, the return on required years of education are necessary as a reference for the returns on overeducation. The additional requirements of a job in terms of education will be compensated with a higher wage.

Gender. Gender is found to be a determinant of wage. Females still experience a wage penalty compared to males.

Experience. Experience is used as a control variable to cover account for some of the unobserved abilities a worker possesses. Additional experience will increase the wage of a worker. The squared experience variable is also included to be able to incorporate diminishing returns as an additional year of experience may not increase wages as much at 20 years of experience compared to 5 years of experience.

3.2. HYPOTHESES

The main research question of this thesis is: *To what extent does overeducation's effect on the wage of a worker vary across occupational groups and industries in the Netherlands?* Thus, this study will test, using Dutch data, if there is a wage penalty compared to properly matched workers with the same education, a wage premium compared to properly matched workers with the same job, and the influence of industry and occupation type on the wage effect. The main research question can be answered by determining if there is an influence caused by the industry and occupation types. To determine this influence, the following hypotheses, corresponding to the red arrows in Figure 3.1, will be tested:

H1a: *A positive relationship exists between the overeducation of a worker and his or her wage compared to properly matched workers with the same required education.*
In the conceptual model of this study (Equation 3.1), this effect is indicated by the returns on overeducation, the coefficient β_2 corresponding to overeducation. If this coefficient is found to be positive and significant, this hypothesis is true.

H2a: *There is a wage penalty for overeducated workers compared to properly matched workers the same obtained education.*
This hypothesis is represented by comparing the coefficient for overeducation, β_2 and required years of education, β_1, in the conceptual model. If the returns on required years of education is bigger than the returns on overeducation, this hypothesis is accepted.

H3a: *The industry a worker is employed in moderates the relationship between overeducation and wage.*
The influence of the industry type is examined for this hypothesis. When the coefficient belonging to the interaction term of a certain industry and overeducation, β_7, exists and is significant, the relationship is indeed moderated.

H4a: *The occupation of a worker moderates the relationship between overeducation and wage.*
The inclusion of the occupation type is examined. This hypothesis investigates if the wage effects differs for specific occupation groups. If it does and is significant, there exists a coefficient term corresponding to the interaction between the occupation group and overeducation, β_8. Then, this hypothesis is accepted.

H3a and H4a are the main hypotheses which will answer the main research question as they cover the wage effect of overeducation and how it differs across occupations and industries. The next section (Section 4) will explain how the hypotheses will be tested.

4

METHODOLOGY

To reach the research objective of this thesis, it is important to think about the methodology of the research. The research objective, initially addressed in Section 1, is restated as it is used as the foundation of the methodology.

> *Research Objective: to investigate and analyze the effect of overeducation on the wage of workers across occupations and industries in the Netherlands*

This section outlines and discusses the research design to achieve the objective. To do so, the formulated sub-research questions are used to dive into the different aspects of this thesis and to be able to, in the end, answer the main research question.

4.1. RESEARCH DESIGN

The types of problems and research questions each require a specific research design to properly examine and answer them. According to Sekaran and Bougie (2016), research questions can be divided into three categories: exploratory, descriptive, and causal. Exploratory research questions are used when existing research is limited, unclear, complex, or there is insufficient material in order create a framework. Often a qualitative approach is used to collect data, such as interviews, focus groups, or case studies. As these researches progress, they become more specific. Furthermore, the findings from exploratory studies are typically not generalizable to the population. On the other hand, when one is interested in understanding characteristics of situations, events, or objects, it is possible to obtain data that describes this topic of interest using a descriptive study. For these studies, one can use either a quantitative or qualitative approach. Lastly, causal research questions are used to examine if a certain variable causes change in a different variable (Sekaran & Bougie, 2016).

As mentioned, the research design should fit the research objective and questions. This research aims to investigate the effect of overeducation on wage. Typically, a project will address multiple types of research questions. Our study also contains different categories of questions:

SRQ1: *How is overeducation estimated?*
This question aims to describe the process used to estimate overeducation. Hence, it is a descriptive research question and will be answered using a qualitative approach. The literature review (Section 2) has already answered this question. The knowledge obtained in the review will be used in the quantitative analysis.

SRQ2: *What variables influence the relationship between overeducation and wage?*
Identifying which variables are important for this research is done through a qualitative approach, as it is an exploratory research question. The literature review (Section 2) has determined these variables and they are incorporated in the conceptual model in Section 3.

SRQ3: *Which industries and occupations are affected the most by overeducation and its effect on wage in the Netherlands?*
This question is a descriptive question as it describes in which context the effect is more prevalent. To answer this question a quantitative design is used, namely a regression analysis.

Main RQ: *To what extent does overeducation's effect on the wage of a worker vary across occupational groups and industries in the Netherlands?*
The main research question is clearly a causal research question. This thesis aims to investigate to what extent a change in overeducation causes a change in wage and how this phenomenon varies across industries and wages. To examine this effect a quantitative approach is used. The conceptual model, developed in Section 3, is used to test the hypotheses and answer this question through regression analysis. A regression analysis is used, because this study aims to examine if there is an effect corresponding to a specific variable. For example, when a t-test is used one can only compare the incidence of variables. With the regression analysis, it is possible to calculate the beta coefficients that are described in Section 3 to test the hypotheses from Section 3.2.

Thus, some questions are already answered in the literature review. By using the knowledge obtained in the review, a theoretical framework is made and used to make hypotheses. These hypotheses will be tested through a regression analysis in order to answer the main research question.

4.2. QUANTITATIVE APPROACH

This study adopts a quantitative research design, which is particularly suitable for examining the causal relationship between overeducation and wages through statistical analysis of numerical data. The quantitative approach takes the form of a regression analysis, which evaluates the causal relationship between an independent variable and a dependent variable, overeducation and wage respectively. Additionally, the impact of the industry and occupation are also examined. It is important to note the distinction between correlation and causation in this approach. Correlation indicates a relationship between two variables, but it doesn't imply one variable causes a change in another one. Sekaran and Bougie (2016) define four conditions to be met for causality:

1. The independent and dependent variable should covary.
2. The independent variable should precede the dependent variable.
3. No other factor should be a possible cause of the change in the dependent variable. This is also why the notion of ceteris paribus is important in econometrics (Wooldridge, 2019). Ceteris paribus means that all other variables are held equal, otherwise one cannot know the causal effect between two variables.
4. A theory is needed and it must explain why the independent variable affects the dependent variable.

Due to the numerous factors that can affect wages, it can be challenging to establish causality with certainty because of the notion of ceteris paribus. This is a prevalent challenge in economic studies, as the data mostly has a nonexperimental nature (Wooldridge, 2019). It is not possible to keep all other factors constant, as is possible in an experiment. However, due to the literature review, which provided a foundation, combined with the regression analysis, at the very least indications of a potential causal relationship can be obtained from this research.

To perform a regression analysis, some steps have to be completed first. First of all, data that represents the population has to be collected and prepared for the analysis. This process is explained further and done in the next section about data, Section 5. This part will also cover how outliers, inconsistencies, and blanks are handled (Sekaran & Bougie, 2016). Then, the next step will be to get a feeling for the data through descriptive statistics. Descriptive statistics are used to summarize the data, providing an overview of the different variables. Afterwards, the data is ready to test the hypothesis. Therefore, the model and hypotheses of Section 3 are used. As mentioned, multiple linear regression analyses are used to test the hypotheses using the created econometric model. The coefficients which are calculated by the regression analysis are examined to determine if the hypotheses can be accepted. All of these steps are performed using statistical software called R (Version 2024.04.1+748).

4.3. REGRESSION ANALYSIS

Having collected, prepared, and explored the data, a regression analysis is used to check for causality and test the hypotheses. Therefore, it is important to understand how regression is done and and how

it is used in this thesis. Let us start with a simple linear regression model (Wooldridge, 2019), shown in Equation 4.1.

$$y = \beta_0 + \beta_1 x + u \tag{4.1}$$

In this case, one wants to study how y varies with changes in x. So, this model contains two variables x and y. u is an error term and contains unobserved factors influencing y. β_0 and β_1, the intercept and slope respectively, are unknown and the goal is to estimate them. Once again, the notion of ceteris paribus is important. If ceteris paribus applies (i.e. all other factors are held constant), then an increase in x by one unit will change y by β_1 units. So, u does not change as x changes. Wooldridge (2019) explains how a method called the Ordinary Least Squares (OLS) can be used to calculate the estimator of the intercept (β_0) slope parameter (β_1). These estimators are denoted as $\hat{\beta}_0$ and $\hat{\beta}_1$. When the estimators are determined, a sample regression function, also known as an OLS regression line, can be written down:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \tag{4.2}$$

In this case Equation 4.2 will be used to predict y for a value of x. How accurate predictions will be is dependent on how closely the estimators represent reality, because the estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ can be biased. For unbiasedness of the estimators, four assumptions are needed, which are part of the Gauss-Markov assumptions for simple regression (Wooldridge, 2019):

1. The population model should be defined as: $y = \beta_0 + \beta_1 x + u$.
2. The sampling from the population should be random.
3. There should be some sample variation in the x.
4. The zero conditional mean assumption should hold, error u has an expected value of zero for any value of x: $E(u|x) = 0$.
5. The error u has the same variance for any value of x: $Var(u|x) = \sigma^2$.

The first four assumptions are needed to show unbiasedness for the estimators. The fifth Gauss-Markov assumption is only used to obtain the variance formulas for the estimators. When assumption 5 is true, it is referred to as homoskedasticity.
Wooldridge (2019) mentions that for econometric studies the key question does remain: have enough other factors been held constant to make a cause for causality? This is related to the fourth assumption. Questions that arise are: what are the omitted variables and are they likely to be correlated with x? If this is the case, assumption 4 will not hold resulting in a biased causal effect β_1. This thesis will look at the effect of overeducation on wage. As described in Section 2, more variables than only the education explain someone's wage and will probably also correlate with overeducation. Hence, multiple x-variables are needed. This is possible in multivariate regression, which uses a multivariate linear regression model with k explanatory variables:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + u \tag{4.3}$$

Similarly as with simple linear regression, multivariate regression uses OLS to estimate the population parameters β_1, ...,β_k. Controlling for more variables can make the conditional mean independence closer to being true. To prove unbiasedness of the multivariate estimators, assumptions are again needed, the Gauss-Markov assumptions for multivariate regression (Wooldridge, 2019):

1. The population model should be defined as: $y = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + u$.
2. The sampling from the population should be random.
3. There should be no perfect collinearity in the sample.
4. The zero conditional mean assumption should hold, error u has an expected value of zero for any value of x: $E(u|x_1, ..., x_k) = 0$.
5. The error u has the same variance for any value of x_k: $Var(u|x_1, ..., x_k) = \sigma^2$.

Compared to the Gauss-Markov assumptions for simple regression, the fifth assumption adds more than only the ability to derive the variance formulas. Again, the first four assumptions are needed for unbiasedness for β_0 and β_1. However, combined with the fifth assumption, it is also determined that the estimators are the best linear unbiased estimators. This means that no other estimators will be better estimators than these estimators which are determined using OLS regression. So, when the assumptions are closer to being true, the estimated coefficients will be more accurate.

In order to conclude something about the population model from a random sample, statistical inference is used. It is important to note that the data for this research is cross-sectional. As a result, any additional treatment given to data containing a time dimension, which often correlates across time, is not necessary. For cross-sectional regression, another assumption is added to the Gauss-Markov assumptions:

6. The error u is independent of the explanatory variables, $x_1, x_2, ..., x_k$ and is normally distributed with zero mean and a constant variance: $u \sim Normal(0, \sigma^2)$

Together these six assumptions form the Classical Linear Model (CLM) assumptions (Wooldridge, 2019). Under these CLM assumptions, the OLS estimators are normally distributed and the t statistics have t distributions under the null hypothesis. Then, the t statistics can be used to test the hypotheses. Hence, it will be discussed how it is possible to check if the CLM assumptions are satisfied to a reasonable degree.

The population model that is going to be used is described by Equation 3.1. This model is a linear combination of the variables. Thus, the population model is defined as stated and this assumption holds.
The sample frame that is used for the Dutch PIAAC data is the 2011 population registry of the Netherlands. From this data a simple random sampling method is used within explicit strata, being municipalities. Because of this sampling method, it is assumed that the sampling from the population can be determined as random and the sample acts as an accurate representation of the population. For the analyses, different subsets of the data for specific industry and occupation types are also examined. Again, it is assumed that it is a random sample, and thus, representative for the specific industry or occupation that is investigated.
To check for multicollinearity two measures can be used, a correlation matrix using Pearson's R and the Variance Inflation Factor (VIF). A correlation matrix, using Pearons's R, for the variables is made and can be seen in Appendix C. Only high values of the coefficient are found between the dummies for overeducation or undereducation and the years of overeducation or undereducation, 0.806 and 0.844 respectively. This is not unexpected as these variables are constructed with the variables for the years of over- and undereducation. However, multicollinearity only problems a major problem in the case of perfect collinearity, a correlation coefficient of 1. This is not the case in this research. Therefore, multicollinearity is not a problem for the Dutch PIAAC data that is included.
The notion of homoskedasticity can be dismissed. There are tests that can determine if homoskedasticity is a proper assumption. Such a test is called the Breusch-Pagan test (Wooldridge, 2019). However, another solution to the problem of possible heteroskedasticity is using heteroskedasticity-robust standard errors. These standard errors do not require CLM assumption 5 to be true. However, problems arise with heteroskedasticity-robust standard errors in small samples, but because of the size of the study's sample, no problems are expected. Therefore, heteroskedasticity-robust standard errors are used in this thesis.

Arguably the most important assumption is the fourth CLM assumption, referring to the zero conditional mean. The assumption is based on the correlation between the error and the explanatory variables. When they are correlated, endogeneity problems occur. This is also illustrated by the literature review, indicating that the omitted variable bias, a source of endogeneity problems, is a big problem. How does this study tackle the problem of the omitted variable bias?
First of all, this study is based on other studies. These studies have examined the effect of overeducation on wage or other determinants of a worker's earnings. To do so, they also had to mitigate the omitted variable bias while identifying these other variables. Combining the research has contributed to a conceptual model, which is described in Section 3. The variables that matter and are necessary to explain wage are included. This enables this research to make better conclusions.
Furthermore, the first step in our regression analysis is to gradually include more variables. When more variables are included, the differences in the statistics of the regression are investigated and it is determined if the additional variable(s) improve the model.

This research will use the statistical software R, in which an OLS regression method is also available through the *lm* function. To create clear output tables of these regressions, the *stargazer* package is used (Hlavac, 2022). These output tables show, besides the coefficients, also the number of observations, the standard error, R^2, adjusted-R^2, and the F-statistic. The number of observations

reflect on the size of the sample. The bigger the sample, the more accurate the estimators will be. As a result, the standard errors, corresponding to the estimated coefficients, will decrease. This will make the results more statistically significant. Responding to this, the significance levels are often lowered. Sekaran and Bougie (2016) define the commonly used significance levels to be 5% and 1%. The data used in this research has a large size. Therefore, the significance level that is used during this research is 99 %. R^2 represents the proportion of the variance in the dependent variable that is explained by the independent variables. This is a goodness-of-fit measure. A higher variance (i.e., closer to 1) means a better fit. However, due to how R^2 is calculated, it will only increase when more variables are added. But, adding variables does not necessarily improve the model because of the possibility of multicollinearity. A better alternative is described by Wooldridge (2019). The adjusted-R^2, which penalizes the number of explanatory variables. Hence, it is able to decrease when variables are added. This measure is often used for and to compare multiple linear regressions. Lastly, the F-statistics indicates if the model does a better job at explaining the dependent variable than a model with no independent variables (Wooldridge, 2019). It tells us if the estimated models matter at all or we are better off with just using the mean of the dependent variable. The F-statistic can also be used to compare two regression models. For this, one model has to be nested within the other model, meaning that one model uses the same variables as the other model, but also adds some more variables to its model. Then, it is possible to determine if the inclusion of the additional variables make the that model perform statistically better than the model with less variables with the F-statistic.

4.4. THE REGRESSION MODEL

This subsection is based on the model created in Section 3, which is a variation of the traditional ORU model. Using this model the following econometric model is created and will be used for the regression analysis (Equation 4.4):

$$\begin{aligned} ln(Wage) = \beta_0 + \beta_1 \cdot ReqYrs + \beta_2 \cdot OEDU\_YRS + \beta_3 \cdot UEDU\_YRS + \beta_4 \cdot Gender + \\ \beta_5 \cdot Exp + \beta_6 \cdot Exp^2 + \beta_7 \cdot OEDU\_YRS \cdot Ind + \beta_8 \cdot OEDU\_YRS \cdot Occ + u \end{aligned} \tag{4.4}$$

Section 5 will explain which variables from the data are used or created for Equation 4.4 . The wage in the above equation appears in the natural logarithmic form. Variables that are solely positive, such as wage, are often not homoskedastic or skewed, which can violate the CLM assumptions. Wooldridge (2019) indicates that when one uses the logarithmic form of these variables, these distributions will look more like a normal distribution, which then satisfies the CLM assumptions more. Furthermore it is important to note that due to the use of the natural log, the interpretation of the change is also different. Now, an increase in a independent variable by one unit will change $wage$ by $100 \cdot \beta_1\%$, it is a change in percentage units.

In the models dummy variables will also be used. Dummy variables are used to distinguish between groups (Wooldridge, 2019). Several groups can be created. These groups are compared to the base group for which no dummy variable is part of the model. Binary values, such as gender, also function as dummy variables. A dummy variable is explained by the following example. Gender will be used as a dummy variable, which can take the value of 0 or 1. Let us assume 0 refers to males and 1 represents females. Then, in the model the reference group are males and the coefficient corresponding to the dummy variable of gender indicates the effect of being a female on, in this case, wage. A variation of the model in Equation 4.4 will also use dummy variables for overeducation and undereducation. Hence, the base reference group are properly matched workers. Dummy variables are also used to represent the different types of industries and occupations. This will be explained further in the section 6.

5

DATA DESCRIPTION AND PREPARATION

This section of the thesis provides an overview of the datasets utilized in the analysis, introducing the sources and variables used. This section aims to define the selection criteria and any preparation steps undertaken to prepare the data for the descriptive statistics and the regression analysis.

5.1. DATA SOURCE

The quantitative part of this thesis will use secondary observational (i.e., passively collected) data. This is data that isn't collected by the researcher and writer of this thesis. The data that will be used is data from the PIAAC survey. PIAAC stand for the Programme for the International Assessment of Adult Competencies (PIAAC) and the survey assesses different characteristics of the respondents, such as education and the level of adult skills (related to literacy, numeracy and problem solving) through a survey (OECD, n.d.). Also the background information of the respondents is gathered. The survey is conducted in 40 countries, resulting in over 200.000 respondents. The Netherlands is among those 40 countries. The data also entails information about different occupations and industries, so it is possible to look at the wage effects across industries and occupational groups and their influences.

5.2. PIAAC VARIABLES

The necessary data that represents the chosen variables is identified through the literature and reasoning. A lot of these variables, representing and influencing a vertical mismatch based on education, are available in the PIAAC data. The variables that are selected from the PIAAC dataset and will be used in the regression models are displayed in Table 5.1. Furthermore, Appendix B contains a codebook in which the possible values and corresponding labels of certain variables are described. Some of the variables that are listed in Table 5.1 are not included in Appendix B. These variables are ratios and are not bounded to a limited set of values. An example of such a variable is wage. The next subsections will explain which variables of the PIAAC data represent the variables that are chosen for the regression model.

5.2.1. DEPENDENT VARIABLE

Wage. The wage includes bonuses and is corrected for purchasing power parity (PPP). PPP is used to enable comparison of the purchasing power of countries. The variable EARNHRBONUSPPP contains this information and will represent wage.

5.2.2. INDEPENDENT VARIABLE

Overeducation (and Undereducation). The main independent variable is the variable that indicates overeducation and undereducation. The literature review has laid out multiple ways of measuring both. Each of these measures have pros and cons. The PIAAC dataset contains information that can be used to compute the subjective measure of over- and undereducation. It will use the variables YRSQUAL, the highest level of educational qualification obtained converted into years (derived from B_Q01A, the highest level of qualification in education according to ISCED 1997), and YRSGET, the self-reported computed years of formal education needed to get the job. YRSGET is subtracted from YRSQUAL. If the outcome is negative, the respondent is undereducated. When the outcome

Table 5.1.: Obtained PIAAC variables

Variable	Description
CNTRYID	Country ID according to ISO 3166
GENDER_R	Respondent's gender
B_Q01a	Highest level of qualification in education, using ISCED 1997
YRSQUAL	Highest level of qualification in obtained education imputed into years of education
YRSGET	Imputed years of formal education needed to get the job, reported by the worker
C_D05	Employment status
C_Q09	Years of paid work during lifetime
EARNHRBONUSPPP	Hourly earnings including bonuses for wage and salary earners, PPP corrected in $US
ISCO1C	Occupational classification of respondent's current job at 1-digit level (ISCO 2008)
ISIC1C	Industry classification of respondent's current job at 1-digit level (ISIC rev 4)

is positive the worker is overeducated. It is also an option that the respondent is neither over- or undereducated. The model based on the traditional ORU model will use the outcome of this extraction as variables for OEDU_YRS and UEDU_YRS. Another set of subjective measures are derived where these variables are converted to dummy variables. Then, it does not take the amount of years of over- or undereducation into account. These variables are indicated as OEDU_DUMMY and UEDU_DUMMY.

Industry. This study examines the influence of the type of industries on overeducation's impact on wage. The PIAAC data also provided information about the industry respondents work in. The codebook in Appendix B can be used to determine which value indicates which industry. Because it is not doable to examine and discuss the effect of every group alone, it is chosen to group them together. The following six groups are prescribed by the International Labour Organization (n.d.):

- Agriculture. Activities that use the natural resources from plants or animals.
- Manufacturing. Creating new product by physically or chemically transforming materials, substances, or components.
- Construction. The activities belonging to the construction of buildings and civil engineering works.
- Mining and quarrying; electricity, gas, and water supply. The extraction of minerals; activities belonging to power, gas, and water supply.
- Market services. Services that generate profit for businesses that provide the service. Examples are transportation, accommodation activities, wholesale.
- Non-market services. Activities that are provided for free or significantly discounted. For example, activites found in the health and social work industry.

Occupational Group. Another effect that is analyzed by this research is the effect of occupational group on the wage effect caused by overeducation. PIAAC gathers information about the occupation of respondents. Again, Appendix B shows the values and corresponding labels of this variable. These

occupation types can be grouped together. An often used grouping divides them into (European Foundation for the Improvement of Living and Working Conditions, n.d.):

- High skilled white collar workers
- Low skilled white collar workers
- High skilled blue collar workers
- Low skilled blue collar workers

The term blue collar refers to jobs that require manual labor, while white collar workers are often represented by office workers. When these groupings are used, the occupation type 'Armed forces' are not assigned to a group as this is a special occupation type. Also, because the Dutch PIAAC data does not contain a lot of observations with armed forces, it is not specifically looked at.

5.2.3. CONTROL VARIABLES

In the literature review different variables were identified that affect the wage of a worker and are correlated with overeducation. These could and should therefore be used as control variables in our regression analysis. The control variables that will be accounted for in the estimation model are: years of education, gender, and experience.

Required years of education. The amount of years in education required to get the job is used as a control variable, RequiredYears. The variable YRSGET is directly usable to represent the amount of years in education that are required to get the job as it indicates the imputed amount of years based on the self-reported educational requirements. This variable will be normalized. By doing so it is prevented that large values of this variable distort the regression.

Gender. The variable GENDER_R indicates the gender of the respondent. Gender will be a dummy variable. Males are used as the reference category.

Experience. Also the years of work experience are included in the model as a control variable. The variable C_Q09 can be used to represent years of work experience as it measured the years of paid work during lifetime. Again, because this variable can take on relatively large values compared to the other variables, it is also standardized.

5.2.4. FILTER VARIABLES

This thesis only focuses on specific data. In order to select this data some variables are needed to filter out irrelevant data.

Country. This research focuses on the Netherlands. The complete PIAAC dataset contains data about multiple countries. Therefore, a variable representing a country's ID is used to filter out any observations that do not correspond to the Netherlands, leaving us with information about only the Dutch respondents.

Employment Status. The data contains information about employed and unemployed people. It is chosen to only use the data of the employed part of the PIAAC dataset. At the same time respondents that are out of the labor force are filtered out. This way the wage can be easily compared without the effect of inflation. The PIAAC variable C_D05 is used for this application, which classifies respondents into the groups employed, unemployed, and out of the labor force.

5.3. DATA PREPARATION

The first step is importing the data. The whole PIAAC dataset has 221308 observations, containing information about 1328 variables. From these, certain variables are selected, as explained in Section 5.2. The scope of this research is the Netherlands. The observations are filtered to Dutch respondents who are employed using the corresponding variables. There are 5170 Dutch respondents and after filtering to currently employed people, the dataset shrinks to from 5170 to 4028 observations. The data contains rows, which do not contain any information about the type of industry or occupational groups. Data may not be known or the respondent did not provide any information about this aspect

of his/her job. These observations are also dropped from the dataset for this research. Lastly, the rows which completely miss any information (i.e., empty cells) will also be dropped. At this point, the dataset still contains 3207 observations. Lastly, the observations corresponding to workers in the armed forces are removed, leaving 3198 observations.

The next step is to identify actions to prepare the data for the analysis such as identifying outliers and handling them or standardizing certain variables. The variables related to the years of experience, years of obtained education, and required years of education are relatively large in terms of size. This can distort the regression results when they are included. In the regressions, only years of experience is used, which have to be standardized beforehand. Furthermore, some variables also have to be transformed, such as gender into a dummy variable, and some variables have to be created, for example variables indicating overeducation. The transformations and creation of variables are described in Tables 5.2 and 5.3 respectively. The variables that are required for the regression models have now been prepared by transforming existing variables or creating new variables using the existing variables. The final list of variables is shown in Table 5.4.

5.3.1. OUTLIER ANALYSIS

After examination of the data, another variable is identified which requires the attention, the wage of the workers. The variable represents the hourly wage of the respondent. The biggest value for this variable is a hourly wage of almost 400.000 US Dollars. This seems unrealistic, so a short outlier analysis is conducted. A boxplot is used in Figure 5.1 to show how much the maximum value of wage differs from the rest of the data points. When this single data point is removed, the mean of wage is reduced by approximately 122 US Dollars to 27.164 US Dollars, which seems to be a more realistic mean hourly wage than 149 US Dollars. However, afterwards there is still one outlier, a hourly wage of around 7628 US Dollars. Removing this observation reduces the mean to 24.807 US Dollars. Hence, these observations are removed of the dataset, afterwards containing 3196 observations.

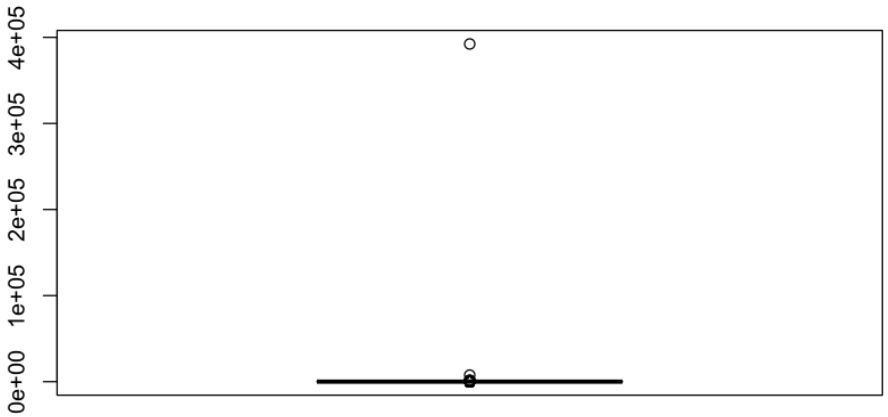

Figure 5.1.: Boxplot of the variable: Wage

Table 5.2.: Transformations performed on variables

Original Variable Name	Transformed to	Transformed by	New Variable Name
GENDER_R	Dummy variable, reference group is males	Takes the value of 0 for males and 1 for females	Gender
C_Q09	Normalized work experience variable	Using the scale() function in R	Experience
EARNHRBONUSPPP	-	-	Wage

Table 5.3.: Creation of new variables explained

New Variable Name	Created by	Possible Values
OEDU_YRS	Equal to YRSQUAL-YRSGET if positive, otherwise equal to 0	Values range from 0 to 14
UEDU_YRS	Equal to the absolute value of YRSQUAL-YRSGET if it is negative, otherwise equal to 0	Values range from 0 to 14
OEDU_DUMMY	Equal to 1 if OEDU_YRS>0, otherwise 0	Dummy variable, 0 or 1
UEDU_DUMMY	Equal to 1 if UEDU_YRS>0, otherwise 0	Dummy variable, 0 or 1
Occupation group name	Dummy made for each group	Value given belonging to new group based on current code
Industry group name	Dummy made for each group	Value given belonging to new group based on current code

Table 5.4.: List of final variables

Variable Name	Variable Type	Variable Unit
Wage	Dependent	US Dollars
OEDU_YRS	Independent	Years
UEDU_YRS	Independent	Years
OEDU_DUMMY	Independent	No unit, dummy variable
UEDU_DUMMY	Independent	No unit, dummy variable
Occupation	Independent	No unit, dummy variable
Industry	Independent	No unit, dummy variable
RequiredYears	Control	Years
Gender	Control	No unit, dummy variable
Experience	Control	No unit, standardized

6

RESULTS

This section contains the results of this research. First the descriptive statistics are examined. Afterwards, the results of the multiple linear regressions are shown. By using this order, this study follows the order of univariate, bivariate, and then multivariate analysis.

6.1. DESCRIPTIVE STATISTICS

Now that the data is prepared for the analysis of this study, the results will begin with the descriptive statistics of the data. Table 6.1 shows the descriptive statistics of the whole Dutch dataset.

Table 6.1.: Descriptive statistic

Statistic	N	Mean	St. Dev.	Min	Max
Gender	3,196	0.501	0.500	0	1
Experience	3,196	0.000	1.000	−1.591	2.565
RequiredYears	3,196	13.471	2.888	7	21
Wage	3,196	24.807	55.096	0.007	1,525.697
OEDU_YRS	3,196	0.713	1.621	0	10
UEDU_YRS	3,196	0.714	1.398	0	9
OEDU_DUMMY	3,196	0.230	0.421	0	1
UEDU_DUMMY	3,196	0.268	0.443	0	1

Table 6.1 can be used to gain some initial insights and the range of the data. Let's start by looking at the mean of some of the variables. Approximately half of the sample is male, and half of the population is female. The mean of the overeducation dummy is equal to 0.23 and of the mean of the undereducation dummy is equal to 0.268. This indicates that 23% of the Dutch sample is overeducated and approximately 27% is undereducated. Furthermore, the wage has a relatively large standard deviation compared to the mean. This shows us that there is a lot of variance in the wage of the observations. Even after removing the extreme values during the data preparation, the wage variable is still spread out over a wide range.

However, this table does not show any information about the occupation or industry variable. The mean of these variable does not contain much information. Instead we will look at the the distribution of the observations across the occupations and industries. As explained in Section 5, some industries and occupations do not contain much observations. Regarding the occupation types, only 20 workers are included of the skilled agricultural and fishery workers occupations. This sample size is too small to make any inference and it is not possible to examine all the different types individually. Therefore,

the occupation and industry types are grouped together. The distribution of the observations across the new groups of occupations and industries are shown in are shown in Tables 6.2 and 6.3. As can be seen, the majority of the workers in the PIAAC data of the Netherlands are white collar workers, slightly above 80% of the sample. This indicates that a lot of the labor in the Netherlands is office work, which also requires a high skill level. Looking at the industry types, the number of observations for the industries related to services are also biggest. This is in line with the finding that most workers are white collar workers as service work is often office work. Although most of the types contain enough information to examine the subset corresponding to the occupation or industry type, this is not the case for the industry type agriculture and mining and quarrying; electricity, gas and water supply. These groups only contain 25 and 31 observations respectively. The amount of variables that are investigated compared to the amount of observations results in not being able to specifically look at the subset of these groups during the regression analysis.

Table 6.2.: Number of observations by occupation type

Occupation type	**Number of observations**
High skilled white collar workers	1620
Low skilled white collar workers	985
High skilled blue collar workers	226
Low skilled blue collar workers	365

Table 6.3.: Number of observations by industry type

Industry type	**Number of observations**
Agriculture	25
Manufacturing	425
Construction	153
Mining and quarrying; Electricity, gas, and water supply (M&Q, EGWS)	31
Market services	1255
Non-market services	1307

Next, let us look at the proportion of overeducation and wage across the industries and occupations. First, the occupation types are examined. Figure 6.1 indicates that three out of the four occupation groups show on average a higher proportion of overeducated workers than the average of the whole sample. This could be because of the fact that the education requirements for these occupation types are lower due to the fact that the level of skills and specialization requirements are lower. Lower requirements increase the chance of being overeducated compared to higher requirements. The average wage across these groups are presented in Figure 6.2. It is shown, as expected, that high skilled white collar workers earn more. For the occupation types, it would be interesting to see if the reduced wage is solely reflected by the higher requirements or also because of the wage penalty for overeducation.

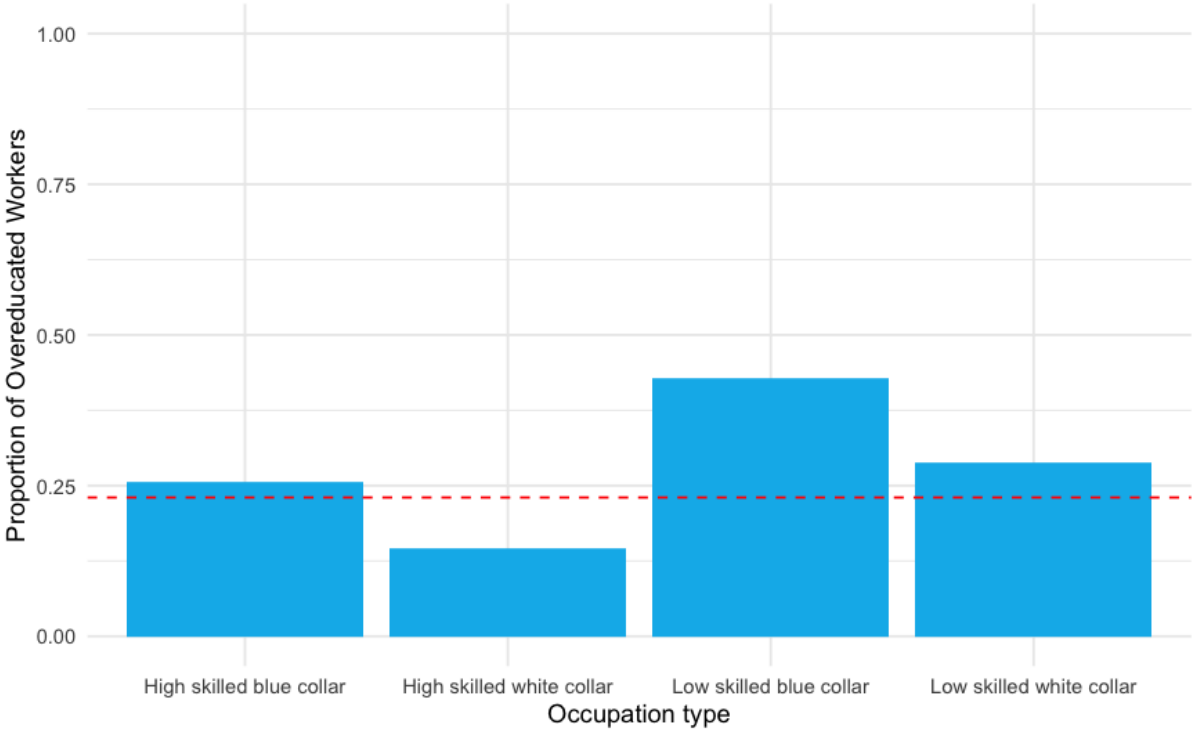

Figure 6.1.: Proportion of overeducated workers by occupation type
Note: The red line represent the proportion of overeducated workers in the whole sample.

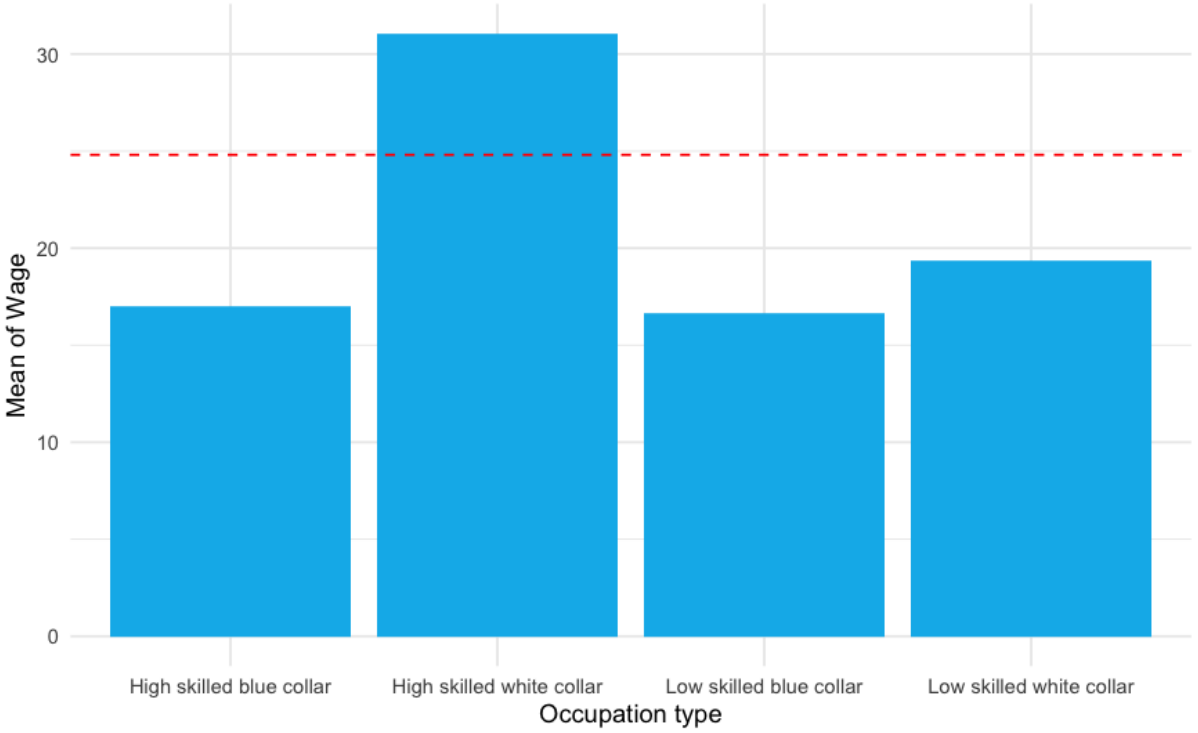

Figure 6.2.: Wage of workers by occupation type
Note: The red line represent the average wage of the whole sample.

The same variables are again examined, but this time looking at the different industry types. Figure 6.3 tells us that the agriculture, M&Q, EGWS, and market services industries have a bigger proportion of overeducated workers compared to the proportion of the whole sample. In Figure 6.4, the average

wage by industry type is shown. Again, the question remains what explains the higher wages.

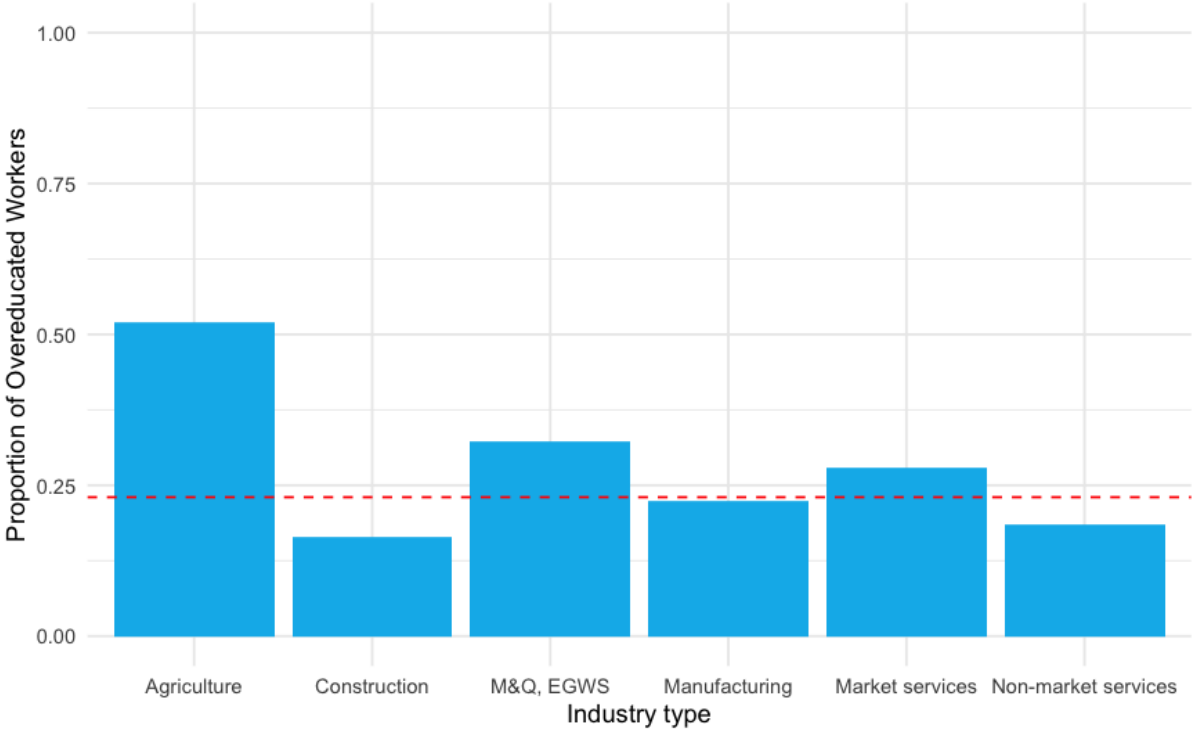

Figure 6.3.: Proportion of overeducated workers by industry type
Note: The red line represent the proportion of overeducated workers in the whole sample.

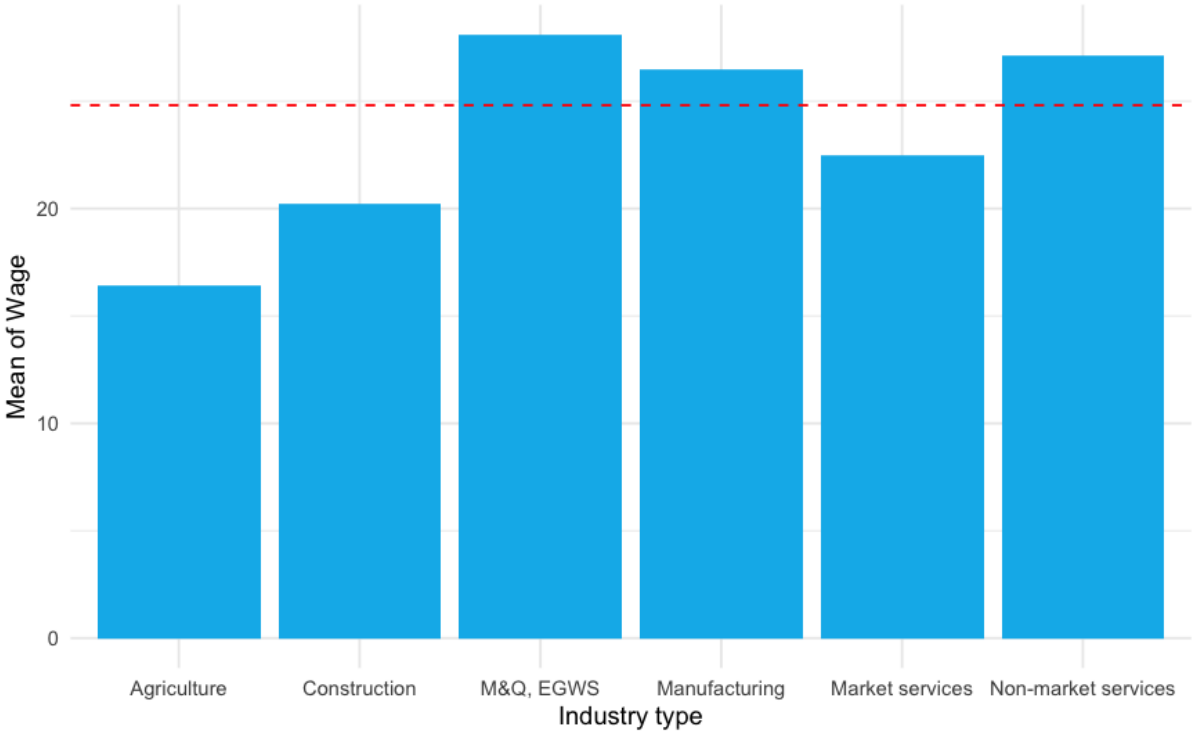

Figure 6.4.: Wage of workers by industry type
Note: The red line represent the average wage of the whole sample.

6.2. MULTIPLE LINEAR REGRESSION

The main focus of this study is to examine the effect of overeducation on wage and how this is influenced by occupation and industry types. This is done through a multiple linear regression analysis, of which the results are shown in this section. The hypotheses that are tested are explained in 3.2. That section also explains when the hypotheses are accepted.

6.2.1. REGRESSIONS WITHOUT INDUSTRY OR OCCUPATION

First, the regression model defined by Equation 4.4 is tested without including variables for the industry and occupation type. The whole sample is used. The results of this are shown in Table 6.4. Table 6.4 is used to check if the base regression model is significant, finds the expected effects, and includes relevant variables that add to the fit of the model to the sample data.

COMPLETE MODEL: THIRD COLUMN

A closer look at the third column is necessary to interpret the complete model. First of all, all of the coefficient in third column are statistically significant at the 0.01 significance level. The coefficients that are the most interesting are bold in Table 6.4. These bold numbers are the coefficients which can be used to test the hypotheses H1a and H2a, the hypotheses related to the wage premium compared to workers with the same required education and the wage penalty compared to workers with the same obtained education respectively.

Returns on over- and undereducation. The returns on overeducation are positive and the returns on undereducation are negative. One additional year of overeducation increases the wage by 5.6%. An additional year of undereducation decreases the wage by 4.8%. The return of required years of education is positive and also bigger than the returns on overeducation. When one more year of education is required, the wage will increase with 11%. These coefficients show that there exists a wage penalty for overeducated workers compared to properly matched workers with the same education, as the return on overeducation is smaller than the return on required years of education. Also, there is a wage premium for overeducated workers compared to properly matched workers with the same job as the returns on overeducation is positive and significant. Therefore, this table provides evidence to support hypotheses H1a and H2a. It is now concluded that there is an effect of overeducation on the wage of a worker for the whole sample, which is in line with the assignment theory as the hypotheses H1a and H2a are based on the. The effect of overeducation that has been found in the literature has been confirmed for the Dutch PIAAC data. The direction of the effects are the same, but the size of the effect is a bit bigger than explained by Leuven and Oosterbeek (2011). The difference between the returns on required years of education is equal to 0.021, between the returns on overeducation 0.013, and between the returns on undereducation 0.012. For all three of these effects, the magnitude is bigger in this study. This is compared to the average estimates of these returns. The returns that are found while using the self-assessment measure are even lower, resulting in even bigger differences in magnitude. If the wage effects that are found are robustly there, it should also visible when subsets of the sample are examined.

Effect of other variables. Besides these focal effects, the returns on control variables are also significant and some interesting observations can be made. Females experience a wage penalty of 9.9%. And an increase of one standard deviation of experience causes wages to increase, but the diminishing returns on experience are reflected by the negative coefficient belonging to the squared experienced variable.

CHECKING FOR RELEVANT VARIABLES: FIRST, SECOND, AND FOURTH COLUMN

To check whether the variables that are included are relevant and improve the fit of the models to the data, the columns are compared. The first column does only include the required years of education and gender of the worker. The effects are significant. The F statistic with a significance level of 0.01 indicates that the model does a better job at explaining the sample than using no independent variable. However, approximately only 23% of the variance is explained by the determinants. When experience is added to the model in the second column, the adjusted-R^2 value increases to 0.345. A F-test where the first column is the restricted model and the second model is the unrestricted model is also significant at the 0.001 level. Hence, the addition of the experience improves the model's

Table 6.4.: MLR of the effect on wage

	Dependent variable: log(Wage)			
	(1)	(2)	(3)	(4)
OEDU_YRS			**0.056*****	
			(0.010)	
UEDU_YRS			**−0.048*****	
			(0.007)	
OEDU_DUMMY				0.125***
				(0.022)
UEDU_DUMMY				−0.123***
				(0.022)
RequiredYears	0.107***	0.086***	**0.110*****	0.099***
	(0.004)	(0.004)	(0.004)	(0.004)
Female	-0.110***	-0.100***	-0.099***	-0.100***
	(0.020)	(0.018)	(0.018)	(0.018)
Experience		0.221***	0.232***	0.230***
		(0.011)	(0.011)	(0.011)
$Experience^2$		−0.129***	−0.115***	−0.121***
		(0.011)	(0.011)	(0.011)
Constant	1.457***	1.880***	1.531***	1.699***
	(0.052)	(0.054)	(0.063)	(0.056)
Observations	3,196	3,196	3,196	3,196
R^2	0.235	0.345	0.369	0.359
Adjusted R^2	0.234	0.345	0.368	0.358
Residual Std. Error	0.570	0.527	0.518	0.522
	(df = 3193)	(df = 3191)	(df = 3189)	(df = 3189)
F Statistic	489.543***	420.977***	311.121***	298.294***
	(df = 2; 3193)	(df = 4; 3191)	(df = 6; 3189)	(df = 6; 3189)
Note:				*p<0.1; **p<0.05; ***p<0.01

fit. The third and fourth model include variables corresponding to overeducation and undereducation. The third column uses years of overeducation and undereducation. The fourth column uses dummies to indicate whether a worker is overeducated or undereducated. The adjusted-R^2 value of the fourth column is lower than the model in the third column. Furthermore, when one compares the F-statistic, while both are statistically significant at the 0.01, the F value of the third column is bigger than the fourth column when they are compared to the model in the second column. Therefore, the model presented in the third column is the model that gives a good fit to the data and has the most explanatory power out of these four models.

6.2.2. REGRESSIONS BY OCCUPATION

The next step is to see if the wage effect is robustly there for each occupation and industry sample. To examine this, the model in the third column of Table 6.4 is used to do regression analyses on samples for each industry and occupation type, as this model provided the best performance. First, the samples belonging to the occupation types are examined. The results of the regressions are found in Table 6.5. The columns represent the high skilled white collar (HSWC) workers, the low skilled white collar (LSWC) workers, the high skilled blue collar (HSBC) workers, and the low skilled blue collar (LSBC) workers respectively.

One can see that the similar coefficients are found for the HSWC workers, the LSWC workers, and LSBC workers. These coefficients are also similar to the coefficients found in the third column of Table 6.4 and support H1a and H2a. Most of these coefficients are also statistically significant at the 0.01 level. However, the model that uses the HSBC sample finds insignificant coefficients for the return on overeducation, undereducation, and required years of education (coefficients in bold). The size of the effects is also completely different than in the other models. Additionally, the explained variance by the variables of these models, R^2, fluctuates quite a bit. In these sample belonging to the HSBC, the adjusted-R^2 is specifically low. Indicating that the variance of the dependent variables is not explained by the independent variables at the same level as in the other samples. This is not necessarily a problem. It gives an indication that other factors affect the wage of worker in this sample containing the HSBC workers. Because these workers experience very different wage effects than expected, it is chosen to include this occupation as a dummy variable while the whole sample is used to examine whether the wage effect of overeducation differs for this occupation.

Table 6.5.: MLR of the effect on wage by occupation type

	Dependent variable: log(Wage)			
	HSWC	LSWC	HSBC	LSBC
OEDU_YRS	0.043	0.037***	**0.013**	0.053***
	(0.033)	(0.011)	(0.026)	(0.012)
UEDU_YRS	−0.041***	−0.046***	**−0.034**	−0.030**
	(0.010)	(0.012)	(0.023)	(0.015)
RequiredYears	0.113***	0.081***	**0.023**	0.076***
	(0.007)	(0.010)	(0.019)	(0.014)
Female	−0.135***	−0.053	−0.131*	−0.102*
	(0.026)	(0.034)	(0.070)	(0.059)
Experience	0.234***	0.247***	0.208***	0.197***
	(0.019)	(0.018)	(0.040)	(0.025)
Experience^2	−0.125***	−0.111***	−0.085***	−0.106***
	(0.019)	(0.020)	(0.023)	(0.024)
Constant	1.671***	1.929***	2.566***	1.932***
	(0.116)	(0.135)	(0.216)	(0.183)
Observations	1,620	985	226	365
R^2	0.245	0.348	0.116	0.329
Adjusted R^2	0.242	0.344	0.092	0.318
Residual Std. Error	0.536 (df = 1613)	0.455 (df = 978)	0.616 (df = 219)	0.474 (df = 358)
F Statistic	87.293***	86.854***	4.790***	29.268***
	(df = 6; 1613)	(df = 6; 978)	(df = 6; 219)	(df = 6; 358)

Note: *p<0.1; **p<0.05; ***p<0.01

6.2.3. REGRESSIONS BY INDUSTRY

The whole sample is also divided in subsets corresponding to the industry types. The agriculture and mining, quarrying, electricity, gas, and water supply industries are not used to do perform a regression analysis as these samples do not contain enough observations to make any inferences. The results of the regressions which used the four remaining industry types are shown in Table 6.6. Again, the regression belonging to three of the samples, agriculture, market, and non-market, find similar coefficients as are found in the third column of Table 6.4 and as expected support H1a and H2a. However, the workers in a certain industry that experience a different wage effect are workers in the construction industry. The returns on overeducation is bigger and comes close to the returns on required years of education. This indicates that in this sample the additional years of education are rewarded similarly to required years of education. The wage penalty for overeducation is small in this sample. The adjusted-R^2 value of the construction industry is also relatively low compared to the other samples. Similar to the HSBC workers, this can show that other variables that are not included do a better job at determining the wage for this sample. To reiterate, this is not a problem as long as these variables are not correlated with overeducation.

Table 6.6.: MLR of the effect on wage by industry type

	Dependent variable: log(Wage)			
	Agriculture	Construction	Market	Non-market
OEDU_YRS	0.052**	0.073**	0.061***	0.048***
	(0.021)	(0.030)	(0.015)	(0.014)
UEDU_YRS	−0.049***	−0.042	−0.052***	−0.041***
	(0.015)	(0.044)	(0.009)	(0.013)
RequiredYears	0.107***	0.081***	0.120***	0.096***
	(0.015)	(0.023)	(0.007)	(0.007)
Female	−0.092**	0.075	−0.152***	−0.088***
	(0.043)	(0.095)	(0.028)	(0.027)
Experience	0.224***	0.319***	0.254***	0.201***
	(0.031)	(0.087)	(0.015)	(0.019)
Experience^2	−0.153***	−0.129***	−0.096***	−0.110***
	(0.042)	(0.045)	(0.014)	(0.018)
Constant	1.713***	1.889***	1.502***	1.837***
	(0.220)	(0.287)	(0.108)	(0.107)
Observations	425	153	1,255	1,307
R^2	0.261	0.176	0.499	0.264
Adjusted R^2	0.251	0.142	0.496	0.260
Residual Std. Error	0.602 (df = 418)	0.721 (df = 146)	0.473 (df = 1248)	0.500 (df = 1300)
F Statistic	24.620***	5.208***	206.869***	77.628***
	(df = 6; 418)	(df = 6; 146)	(df = 6; 1248)	(df = 6; 1300)

Note: *p<0.1; **p<0.05; ***p<0.01

Due to the large difference compared to what is expected, the construction industry should also be used as a dummy variables to examine whether this industry experiences different wage effect of overeducation compared to the rest of the whole sample. The restricted samples cannot be compared. The samples are different, so one cannot check if the differences are statistically significant between the samples. Hence, no conclusions can be drawn about the differences between the effect of two subsets of the sample. However, it does provide insights in occupation and industries where the effect is different than expected. The high skilled blue collar workers and workers in the construction industry experienced different effects than one would have expected if the same wage effect of overeducation would have been robustly present in the Dutch PIAAC data. In the end, this study aims to say something about the Dutch population and the influence of industry and occupation type on the wage effect of overeducation. Therefore, if one wants to say something about these effects and the influence of these industry and occupation groups in the Dutch labor market, a representative sample is needed, the whole sample.

6.2.4. REGRESSIONS USING DUMMIES FOR OCCUPATION AND INDUSTRY

As explained in the previous two sections, the HSBC workers and workers in the construction industry seem to experience different wage effects than expected. To make any inference about the differences between these groups and the rest of the whole sample, dummies are included for the high skilled blue collar workers and the construction industry. For the next regression the whole sample of the Dutch PIAAC data is used. The dummies are added one at the time to check if the inclusion of the variables results in a better performance of the model. The results are found in Table 6.7.

COMPLETE MODEL INCLUDING BOTH DUMMIES: FOURTH COLUMN

The fourth column is the most important to interpret, as it is the complete model which also includes dummies for the construction term and high skilled blue collar workers. The interaction term of these groups with overeducation is also included as the coefficient belonging to these terms are examined to test hypothesis H3a and H4a, which investigate the influence an industry or occupation has on the wage effect of overeducation.

Returns on over- and undereducation. First of all, the returns on overeducation and required years of education (first two bold coefficients in Table 6.7) are again significant and similar to the coefficients found in Table 6.4. Therefore, H1a and H2a are again accepted. Thus, the coefficients are in line with the assignment theory as the wage of a worker is influenced by the quality of the match based on education. An additional year of overeducation increases the wage with 4.9%. Required years of education have a bigger impact on the wage of a worker as is expected. One additional year that is required for the job increase the wage by 10.7%.

Effect of interaction terms. To test the hypotheses H3a and H4a, the interaction terms are examined. The coefficients corresponding to the interaction terms (last two bold coefficients in Table 6.7) do not show significant effects individually. Although insignificant, the size and direction of the coefficient indicates that for both the workers in the construction industry and high skilled blue collar workers the wage penalty for overeducation decreases and the wage premium increases. In the case that a worker is in the construction industry, the return on a year of overeducation increases from 4.9% to 7.6%. When a worker is a high skilled blue collar worker, the wage effect of overeducation increases from 4.9% to 9%. In these cases, the coefficients indicate that the hypotheses H1a and H2a are still supported when these other variables are included. The wage premium compared to properly matched workers with the same required education still exists. Also, the wage penalty compared to properly matched workers with the same obtained education is still visible as the returns on a year of overeducation remains smaller than the returns on a year of required education. Thus, the coefficients are then still in line with the assignment theory, as explained in Section 2.4.2. H3a and H4a are represented by the coefficients belonging to the interaction terms. The insignificance of the coefficients provides no evidence to reject the null hypotheses and accept H3a and H4a. However, the significant F-test does show that the inclusion helps in explaining the dependent variable. Therefore, more research is needed examine the effect of the occupation and industry types on the wage effect of overeducation.

CHECKING FOR RELEVANT VARIABLES

To check whether the inclusion of the dummy variables improves the model, the dummies are added one at the time. The adjusted-R^2 value of the models only slightly increases. However, when the models are compared to each other using a F-test, it is shown that the model using both dummies fits the data significantly better at the 0.001 level. This indicates that, collectively, the predictors of the fourth model do a better job to fit the whole sample. Hence, the inclusion of the dummies is beneficial and the model in the fourth column should be used to make any inferences about the whole sample.

Table 6.7.: MLR of the effect on wage using dummies for industry and occupation types

	Dependent variable: log(Wage)			
	(1)	(2)	(3)	(4)
OEDU_YRS	0.056***	0.053***	0.050	**0.049***
	(0.010)	(0.010)		(0.010)
UEDU_YRS	−0.048***	−0.048***	−0.048	−0.048***
	(0.007)	(0.007)		(0.007)
RequiredYears	0.110***	0.109***	0.108***	**0.107***
	(0.004)	(0.004)	(0.004)	(0.004)
Female	−0.099***	−0.107***	−0.115***	−0.118***
	(0.018)	(0.018)	(0.018)	(0.018)
Experience	0.232***	0.232***	0.232***	0.232***
	(0.011)	(0.011)	(0.011)	(0.011)
$Experience^2$	−0.115***	−0.115***	−0.114***	−0.114***
	(0.011)	(0.011)	(0.011)	(0.011)
Construction		−0.118*		−0.064
		(0.066)		(0.058)
OEDU_YRS*Construction		0.046*		**0.027**
		(0.026)		(0.033)
HSBC			−0.168	−0.151***
				(0.043)
OEDU_YRS*HSBC			0.048	**0.041**
				(0.027)
Constant	1.630***	1.652***	1.686***	1.692***
	(0.065)	(0.066)	(0.057)	(0.068)
Observations	3,196	3,196	3,196	3,196
R^2	0.369	0.371	0.373	0.373
Adjusted R^2	0.368	0.369	0.371	0.371
Residual Std. Error	0.518	0.517	0.517	0.517
	(df = 3189)	(df = 3187)	(df = 3187)	(df = 3185)
F Statistic	311.121***	234.504***	236.643***	189.474***
	(df = 6; 3189)	(df = 8; 3187)	(df = 8; 3187)	(df = 10; 3185)
Note:				*p<0.1; **p<0.05; ***p<0.01

7

DISCUSSION

This section discusses the results of this study. This study had one research objective: to investigate and analyze the effect of overeducation on the wage of workers across occupations and industries in the Netherlands. This section will discuss and interpret the key findings. Also, the theoretical contributions and implications are covered. Lastly, the limitations of the research are explained and future research recommendations are described.

7.1. KEY FINDINGS

The key findings of this thesis are discussed in the order of the research questions. The hypotheses are restated for clarity.

> **H1a:** *A positive relationship exists between the overeducation of a worker and his or her wage compared to properly matched workers with the same required education.*
>
> **H2a:** *There is a wage penalty for overeducated workers compared to properly matched workers the same obtained education.*
>
> **H3a:** *The industry a worker is employed in moderates the relationship between overeducation and wage.*
>
> **H4a:** *The occupation of a worker moderates the relationship between overeducation and wage.*
>
> **SRQ1:** *How is overeducation defined and measured?*

A literature review was performed as the foundation of this thesis. This section provided information on overeducation and its consequences for wages. It quickly became clear that overeducation consistently results in a wage premium compared to properly matched workers with similar jobs and in a wage penalty compared to employees with the same level of obtained education. Several studies during the period ranging from 1980 till now have found this same result. However, one major problem was identified in the field of educational mismatch, the omitted variable bias. This bias is caused by variables that are correlated with both overeducation and wage but are not included in the model. The studies have (tried to) identify multiple variables to reduce the negative impact of unobserved heterogeneity on regressions. Although many variables are now commonly used in research investigating overeducation's effect on wage, the influence of industry types and overeducation has not been examined yet. Therefore, this thesis has tried to analyze the effect the type of industry or occupation has on the relationship between overeducation and wage. To do so, the first step was to address the question: how does one measure overeducation? It was found that three measures are used in educational mismatch studies, the self-reported measure, the realized-matches measure, and the job evaluation method. Unfortunately, no consensus has been reached on which measure is best used. The use of a certain measure is mostly dependent on the data that is available. When a measure is chosen, it is important to keep in mind the implications of using that specific measure. In the end, this thesis used the subjective measure. This answers sub-research question 1.

> **SRQ2:** *What variables influence the relationship between overeducation and wage?*

Briefly the importance of including (control) variables is mentioned when the omitted variable bias is discussed. This is also an important aspect as this thesis aims to analyze the relationship between an independent variable and dependent variable. In the studies regarding overeducation various variables have been included in the regression models. Theories describe different determinants to be important when looking at wage determination. For some, the determinants are related to the supply side (human capital theory), others use variables related to the demand side (job assignment theory), or a combination of the demand and supply side variables are used to determine the quality of the match between the worker and the job requirements (assignment theory). To determine which theory best represents reality a regression analysis is used. The regression models are often based on the ORU model. The variation of ORU model that is used in this thesis incorporated required years of education, years of overeducation, years of undereducation, work experience, and gender. It is found that the hypotheses H1a and H2a are supported by the evidence and the coefficients are in line with the assignment theory (Sattinger, 1993), which is explained in Section 2.4.2. Using the data under consideration, it is determined that the assignment theory is true. The quality of the match, using variables of both the supply and demand side of the labor market, determines the wage. If the education is not efficiently used, a proper match, the returns on education are in general lower.

> **SRQ3:** *Which industries and occupations are affected the most by overeducation and its effect on wage in the Netherlands?*

It is determined that the influence of specific occupation type or industry type are not examined enough. After examination of the model by industry and occupation type, two types are highlighted that show different wage effects than is expected, the construction industry and the high skilled blue collar workers. Although the inclusion of the interaction terms are beneficial to fit the model to the data, no evidence has been found to accept H3a and H4a as the coefficient of the interaction terms, representing the influence of the industry or occupation type on the wage effect, are insignificant. Still, the study identified that the impact of overeducation on wages varies across different industries and occupational groups. High-skilled blue-collar workers and workers in the construction industry seem to experience a smaller wage penalty compared to other industries and occupation. Indicating that the mismatch between education and job requirements may be less problematic and/or better managed in these areas. The assignment theory by Sattinger (1993) still holds but the size of the wage effect seems to differ. This finding highlights the importance of studies that further investigate the effect of occupation types and industry types on the wage effect caused by overeducation.

> **Main RQ:** *To what extent does overeducation's effect on the wage of a worker vary across occupational groups and industries in the Netherlands?*

Combining all this information will provide answers to the main research question. It is shown that the high skilled blue workers and the construction industries seem to show different effects on the wage of a worker by overeducation. The significant F-statistic when the interaction terms are included show that the inclusion of the occupational group and industry does explain the data better. However, the null hypotheses corresponding to H3a and H4a cannot be rejected due to the insignificant coefficients. Hence, using this particular dataset, no conclusions can be drawn about how much the wage effect of overeducation varies across occupational groups and industries in the Netherlands. This answers the main research question.

7.2. THEORETICAL CONTRIBUTIONS

The issue of overeducation is a problem that has been well-studied in the recent years. Especially since policies in a lot of countries are aimed towards increasing the level of education of the population in recent years due to the changing labor environment. The effect of overeducation on wage is also thoroughly examined in literature. However, the literature review has identified that the research on the influence of specific industries or occupational groups on the effect of overeducation on wage has been limited. Additionally, no studies were found that looked at the problem of overeducation's effect on wage, zooming in on the Netherlands while using the PIAAC data. This study fills this knowledge gap by using PIAAC data of the Netherlands for regression analyses. The wage premium, compared to properly matched workers with the same job, and the wage penalty, compared to properly matched workers with the same education, are also found in this research. Both these findings support the assignment theory, as the statistically significant coefficients are found for the Dutch PIAAC data that

confirm this theory. Therefore, the knowledge in the current literature is validated. Also, this research provides some deeper insights on the effect of certain occupation and industry type on the wage effect of overeducation. It is shown that workers in the construction industry and high skilled blue collar worker seem to experience a smaller wage penalty than the other workers in the whole sample. The assignment theory still holds for workers in this industry or with this occupation, but the size of the effect is different. The wage penalty is smaller and the premium is bigger. This is unexpected. It was expected that the high skilled blue collar jobs, such as technicians or operators, will gain more wage because of the skills that are learned through experience. Then, the years of (over)education should not receive a smaller wage penalty than the other groups. The same holds for the construction industry. An explanation might be given by the fact that there is a shortage of high skilled blue collar workers, tradespersons, in the Netherlands as is explained by NOS (2023). As a results, companies might be trying to tempt workers, with a higher education than necessary, using a higher wage than is normally used for these positions. As a result, the years of overeducation aren't penalized as much. However, more research is necessary to confirm this explanation as this study has not found any significant coefficients to accept hypotheses H3a and H4a, but only found that the inclusion of these types improves the fit of the model to the data.

7.3. PRACTICAL IMPLICATIONS

The introduction (Section 1) has briefly described the challenges due to overeducation on itself and its effect on wage. This research has provided policymakers with insights in problem areas in the Dutch labor market. The influence of a specific industry and occupation that a worker belongs to on the effect of overeducation has been examined. This information can be used to develop fair wage policies in industries and for occupational groups where overeducation impacts wages. Also, policies can be designed to better align the obtained education of the worker with the required education of the worker. For example, educational institutions can adjust their education to provide a better fit to the job.

Another way of addressing the issue of overeducation and its impact on wage is through career guidance services in school. These services can use the information provided by this research to strengthen their claims. By doing so, students and/or workers obtain relevant information when they have to choose their study field. It informs them if the work they aspire to perform requires a specific education and if overeducation is penalized. They can then use this information to choose if the additional education is necessary in their opinion.

7.4. LIMITATIONS AND FUTURE RESEARCH

This research, like many others, has to deal with some limitations. The limitations are discussed and future research recommendations are given to tackle them. First of all, this research has used a measure which uses self-reporting. Inherently self-reporting results in bias, workers will often exaggerate the requirements that are necessary for their job. Additionally the question arises: how good can a worker determine the requirements for his own job? Hence, the reliance on self-reporting measures leads to bias and limits the objectivity of the data. Future research should also use other or improve measures and use other data sources which might be more objective. However, a consensus on which of the measures, the three described in the literature review, is the 'best' is still missing. Hence, at this point in time, checking these results using the other measures will strengthen the findings of this research. Furthermore, improving the measurement of educational requirements would also allow for more accurate results which will improve the validity of the results. Moreover, when a consensus is reached on which measure is the best and will be used, studies can easily be compared.

The scope of the research on wage effect of overeducation and the influence of the industry and occupation type on this effect is limited to the Netherlands. Some other factors might influence the relationship between overeducation and wage, which are not observed in this dataset. For example, the regulatory system might differ per country and result in a different size of the impact. Hence, a recommendation for future research is to also look at the influence of industries and occupations on overeducation's impact on wage in different regions. This will improve the generalizability of the study's findings and create more robust results. The PIAAC only takes into account OECD nations. But there is a big difference between OECD nations and non-OECD nations due to the differences in

regulatory environments. Thus, the importance of broadening the geographical focus is important in order to improve generalizability.

This research finds significant coefficients between the variables in the regression models. However, due to the non-experimental nature causation cannot be determined. Unfortunately this problem is not able to be investigated in a experiment. Therefore, more studies should also examine the impact of industries and occupational groups on overeducation's impact to build up a body of knowledge. Preferably these studies use different data sources and also, as explained, cover different countries. In the end, this combined can strengthen (or weaken) the conclusion which are drawn in this study and add evindence that support the hypotheses.

A major ongoing problem in this part of labor economics and econometric studies is the omitted variable bias. Have enough variables been taken into account that affect the dependent variable? Due to the limited time of this study and available data, only the most used variables are used as control variables, namely gender, experience, and required education. Future research should examine if there are more variables that have a big impact on wage and are not included in this study. These should then be included to improve the wage determination model and consequently be checked if similar findings as this study's results are found. In addition, this study uses cross-sectional data. Time effects could not be observed due to the nature of the data. Therefore, future studies should also examine this topic using data which spans across time.

As mentioned in the introduction, labor is changing. Changes in technology, demographics, policies and globalization lead to the quickly evolving landscape of labor. This thesis uses PIAAC data of the Netherlands which was gathered in 2011-2012. Hence, it would be beneficial to try examine the effect with more recent data. Do the same findings still hold or are other effects found? Later this year (2024) the second cycle of the PIAAC survey is expected to be published. The data of this second cycle can be used to perform this research.

Finally, the results provide an indication of areas where overeducation's effect on wage is more prevalent or not, which policymakers need to address primarily. However, this thesis did not address the specific policies that may be effective to mitigate these problems. More research may be necessary corresponding to the mechanisms causing employees to take jobs below their level of education.

8

CONCLUSION

This thesis examined the effect of overeducation on wage across different industries and occupations, specifically focusing on the Netherlands. The main research question that has been answered is:

> **Main RQ:** *To what extent does overeducation's effect on the wage of a worker vary across occupational groups and industries in the Netherlands?*

The first step in answering this question was a literature review. The literature explains how overeducation is measured, how and to what extent it influences the wage of a worker, but not a single research covers the influence of industries and occupational groups on overeducation's effect on wage. Furthermore, the literature review created a foundation of knowledge that helped to create a theoretical framework with a conceptual model. Ultimately, hypotheses were made representing the relationships between overeducation, wage, industry type, and occupational group.

Then, this study examined the effect of overeducation on wages across different industries and occupations in the Netherlands. The findings confirm that overeducation results in a wage premium compared to properly matched workers in similar jobs but a wage penalty when compared to workers with the same level of education but they are properly matched. The study contributes to the theoretical literature by validating the assignment theory and finding similar coefficients as are found in the literature. It also highlights the need to perform a more detailed examination of the influence of the occupation and industry types on the wage effect of overeducation. Using the Dutch data, it is shown that workers in the construction industry and high skilled blue collar workers seem to experience a smaller wage penalty caused by overeducation compared to properly matched workers with the same obtained education. Hence, occupation and industry type can have an effect on the relationship between overeducation and wage, but more research is necessary to confirm the moderating effect of occupational groups and industries.

Practically, it offers relevant information for policymakers, individuals, education institutes, and employers to develop strategies that mitigate the negative effects of overeducation, improve the efficiency of the labor market, and improve the returns on the public investment in education. Future research should address the limitations of this study by using longitudinal data, exploring additional variables, and conducting comparative analyses across different countries. By doing so, it will aid in developing a more comprehensive and realistic understanding of overeducation's impact on wages and the influence of industry and occupation types. Then, these insights can be used to create more effective policy and educational interventions.

BIBLIOGRAPHY

Allen, J., & van der Velden, R. (2001). Educational mismatches versus skill mismatches: Effects on wages, job satisfaction, and on-the-job search. *Oxford Economic Papers*, *53*(3), 434–452. https://doi.org/10.1093/oep/53.3.434

Becker, G. S. (2009). *Human capital: A theoretical and empirical analysis, with special reference to education*. University of Chicago press.

Brunello, G., & Wruuck, P. (2021). Skill shortages and skill mismatch: A review of the literature. *Journal of Economic Surveys*, *35*(4), 1145–1167. https://doi.org/10.1111/joes.12424

Carroll, D., & Tani, M. (2013). Over-education of recent higher education graduates: New australian panel evidence. *Economics of Education Review*, *32*, 207–218.

CBS. (2022). Meer hoogopgeleiden en beroepsniveau steeg mee. https://www.cbs.nl/nl-nl/nieuws/2022/42/meer-hoogopgeleiden-en-beroepsniveau-steeg-mee#:~:text=%27.,ruim%201%20op%20de%203.

Chłoń-Domińczak, A., & Żurawski, A. (2017). Measuring skills mismatches revisited - introducing sectoral approach.

Choi, A., Guio, J., & Escardibul, J.-O. (2020). The challenge of mapping overeducation and overskilling across countries: A critical approach using piaac. *Compare: A Journal of Comparative and International Education*, *50*(2), 237–256. https://doi.org/10.1080/03057925.2019.1600400

De Santis, M. O., Gáname, M. C., & Moncarz, P. E. (2022). The impact of overeducation on wages of recent economic sciences graduates. *Social Indicators Research*, *163*(1), 409–445.

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

European Foundation for the Improvement of Living and Working Conditions. (n.d.). Coding and classification standards. https://www.eurofound.europa.eu/en/coding-and-classification-standards-0

European Training Foundation. (2019). *Skills mismatch measurement in ETF partner countries.* Publications Office. https://data.europa.eu/doi/10.2816/664496

Groeneveld, S., & Hartog, J. (2004). Overeducation, wages and promotions within the firm. *Labour Economics*, *11*(6), 701–714. https://doi.org/10.1016/j.labeco.2003.11.005

Handel, M. J. (2020). Job Skill Requirements: Levels and Trends.

Hartog, J. (2000). Over-education and earnings: Where are we, where should we go? *Economics of Education Review*, *19*(2), 131–147. https://doi.org/10.1016/S0272-7757(99)00050-3

Hlavac, M. (2022). *Stargazer: Well-formatted regression and summary statistics tables* [R package version 5.2.3]. Social Policy Institute. Bratislava, Slovakia. https://CRAN.R-project.org/package=stargazer

International Labour Organization. (n.d.). The International Standard Classification of Occupations-ISCO-08. https://isco-ilo.netlify.app/en/isco-08/

Iriondo, I., & Pérez-Amaral, T. (2013). The effect of educational mismatch on wages using european panel data.

Korpi, T., & Tåhlin, M. (2009). Educational mismatch, wages, and wage growth: Overeducation in sweden, 1974–2000. *Labour Economics*, *16*(2), 183–193. https://doi.org/10.1016/j.labeco.2008.08.004

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

Levels, M., van der Velden, R., & Allen, J. (2014). Educational mismatches and skills: New empirical tests of old hypotheses. *Oxford Economic Papers*, *66*(4), 959–982. https://doi.org/10.1093/oep/gpu024

Lucas, J. (2022). Aantal bollebozen stijgt tot recordhoogte. https://www.bnr.nl/nieuws/onderwijs/10491529/aantal-bollebozen-stijgt-tot-recordhoogte

Mateos Romero, L., Murillo Huertas, I. P., & Salinas Jiménez, M. D. M. (2017). Wage effects of cognitive skills and educational mismatch in europe. *Journal of Policy Modeling*, *39*(5), 909–927. https://doi.org/10.1016/j.jpolmod.2017.08.001

Mateos-Romero, L., & Salinas-Jiménez, M. d. M. (2018). Labor mismatches: Effects on wages and on job satisfaction in 17 oecd countries. *Social Indicators Research*, *140*(1), 369–391. https://doi.org/10.1007/s11205-017-1830-y

McGowan, M. A., & Andrews, D. (2015). Skill Mismatch and Public Policy in OECD Countries [Type: Journal Article]. https://doi.org/doi:https://doi.org/10.1787/5js1pzw9lnwk-en

McGuinness, S. (2006). Overeducation in the labour market. *Journal of economic surveys*, *20*(3), 387–418.

McGuinness, S., Bergin, A., & Whelan, A. (2018). Overeducation in europe: Trends, convergence, and drivers. *Oxford Economic Papers*, *70*(4), 994–1015.

McGuinness, S., Pouliakas, K., & Redmond, P. (2018). Skills Mismatch: Concepts, Measurement and Policy Approaches. *Journal of Economic Surveys*, *32*(4), 985–1015. https://doi.org/10.1111/joes.12254

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

Nieto, S., & Ramos, R. (2017). Overeducation, skills and wage penalty: Evidence for spain using piaac data. *Social Indicators Research*, *134*(1), 219–236. https://doi.org/10.1007/s11205-016-1423-1

NOS. (2023, June). Groot tekort aan vakmensen, opleidingen naarstig op zoek naar studenten. https://nos.nl/regio/zh-west/artikel/401816-groot-tekort-aan-vakmensen-opleidingen-naarstig-op-zoek-naar-studenten

OECD. (2017). *Getting Skills Right: Skills for Jobs Indicators*. https://doi.org/https://doi.org/https://doi.org/10.1787/9789264277878-en

OECD. (n.d.). Survey of Adult Skills (PIAAC) - PIAAC, the OECD's programme of assessment and analysis of adult skills. https://www.oecd.org/skills/piaac/

Quintini, G. (2011). Right for the job: Over-qualified or under-skilled?

Rubb, S. (2006). Educational mismatches and earnings: Extensions of occupational mobility theory and evidence of human capital depreciation. *Education Economics*, *14*(2), 135–154. https://doi.org/10.1080/09645290600622905

Sattinger, M. (1993). Assignment models of the distribution of earnings. *Journal of economic literature*, *31*(2), 831–880.

Sekaran, U., & Bougie, R. (2016). *Research methods for business: a skill-building approach* (7th ed.). http://103.227.140.9/handle/123456789/18302

Sellami, S., Verhaest, D., Nonneman, W., & Trier, W. V. (2017). The impact of educational mismatches on wages: The influence of measurement error and unobserved heterogeneity. *The B.E. Journal of Economic Analysis Policy*, *17*(1). https://doi.org/10.1515/bejeap-2016-0055

Shahidan, A., & Ismail, R. (2021). A critical review of the literature on the concept of job mismatch and overeducation [Type: Journal Article]. *Journal of Economics and Sustainability*, *3*(1), 9–9.

Thurow, L. C. (1975). Generating inequality.

United Nations Statistics Division. (2008, August). *International Standard Industrial Classification of All Economic Activities (ISIC), Rev.4*. https://doi.org/10.18356/8722852c-en

van der Mooren, F., & de Vries, R. (2022, October). Steeds meer hoogopgeleiden in Nederland: wat voor beroep hebben ze? https://www.cbs.nl/nl-nl/longread/statistische-trends/2022/steeds-meer-hoogopgeleiden-in-nederland-wat-voor-beroep-hebben-ze-?onepage=true

Vandeplas, A., & Thum-Thysen, A. (2019). *Skills mismatch & productivity in the eu*. Publications Office of the European Union Luxembourg.

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

Verhaest, D., & Omey, E. (2006). The impact of overeducation and its measurement. *Social Indicators Research*, *77*(3), 419–448. https://doi.org/10.1007/s11205-005-4276-6

Wooldridge, J. M. (2019, January). *Introductory Econometrics: a Modern Approach*. Cengage Learning.

Yeo, J. Z., & Maani, S. A. (2017). Educational mismatches and earnings in the new zealand labour market. *New Zealand Economic Papers*, *51*(1), 28–48.

A

LITERATURE REVIEW METHODOLOGY

The literature review was conducted using a systematic approach. Search databases such as Google Scholar and Scopus were used to find literature on this research topic. Table A.1 shows the keywords and synonyms that are used as search terms.

Table A.1.: Keywords and Synonyms used for Literature Search

Keywords	**Synonyms**
Overeducation	Occupational mismatch, overqualification, educational mismatch, surplus education
Wage	Earnings, pay
Industry	Sector, trade, field, line of work
Occupation	Occupational group, job, profession, trade, position

The abstracts of the found literature are inspected to determine if the specific literature has any relevance to this study. In order to limit the amount of literature which does not contain any additional value to this research, some criterion are created. The criterion are shown in Table A.2 and used to include or exclude certain works. The literature thas was found to be relevant was also used to find other relevant articles. The references were used for chain searching, which refers to identifying and examining of cited works.

Table A.2.: Criterion for Inclusion and Exclusion

Criterion	**Inclusion**	**Exclusion**
Language	Dutch, English	Non-Dutch, non-English
Geography	Worldwide, preference for Western countries, focus on the Netherlands	None
Citations	Preference for more than 25 citations	Less than 10 citations, unless recent article/paper
Date	1950-now	Before 1950

B

PIAAC DATA: VALUES OF VARIABLES EXPLAINED

Table B.1.: PIAAC Variable with corresponding values and value labels

Variable Name	Value	Value Label
CNTRYID	36	Australia
CNTRYID	40	Austria
CNTRYID	56	Belgium
CNTRYID	124	Canada
CNTRYID	152	Chile
CNTRYID	196	Cyprus
CNTRYID	203	Czech Republic
CNTRYID	208	Denmark
CNTRYID	218	Ecuador
CNTRYID	220	England
CNTRYID	233	Estonia
CNTRYID	246	Finland
CNTRYID	250	France
CNTRYID	276	Germany
CNTRYID	300	Greece
CNTRYID	348	Hungary
CNTRYID	372	Ireland
CNTRYID	376	Israel
CNTRYID	380	Italy
CNTRYID	392	Japan
CNTRYID	398	Kazakhstan
CNTRYID	410	Korea
CNTRYID	440	Lithuania
CNTRYID	484	Mexico

Variable Name	Value	Value Label
CNTRYID	528	Netherlands
CNTRYID	554	New Zealand
CNTRYID	578	Norway
CNTRYID	604	Peru
CNTRYID	616	Poland
CNTRYID	643	Russian Federation
CNTRYID	702	Singapore
CNTRYID	703	Slovak Republic
CNTRYID	705	Slovenia
CNTRYID	724	Spain
CNTRYID	752	Sweden
CNTRYID	792	Turkey
CNTRYID	826	United Kingdom
CNTRYID	840	United States
GENDER_R	1	Male
GENDER_R	2	Female
GENDER_R	9	Not stated or inferred
B_Q01a	1	No formal qualification or below ISCED 1
B_Q01a	2	ISCED 1
B_Q01a	3	ISCED 2
B_Q01a	4	ISCED 3C shorter than 2 years
B_Q01a	5	ISCED 3C 2 years or more
B_Q01a	6	ISCED 3A-B
B_Q01a	7	ISCED 3 (without distinction A-B-C, 2y+)
B_Q01a	8	ISCED 4C
B_Q01a	9	ISCED 4A-B
B_Q01a	10	ISCED 4 (without distinction A-B-C)
B_Q01a	11	ISCED 5B
B_Q01a	12	ISCED 5A, bachelor degree
B_Q01a	13	ISCED 5A, master degree
B_Q01a	14	ISCED 6
B_Q01a	15	Foreign qualification
B_Q01a	16	ISCED 5A bachelor degree, 5A master degree, and 6 (without distinction)
B_Q01a	96	Valid skip
B_Q01a	97	Don't know

Variable Name	Value	Value Label
B_Q01a	98	Refused
B_Q01a	99	Not stated or inferred
C_D05	1	Employed
C_D05	2	Unemployed
C_D05	3	Out of the labour force
C_D05	4	Not known
C_D05	6	Valid skip
C_D05	7	Don't know
C_D05	8	Refused
C_D05	9	Not stated or inferred
C_Q07	1	Full-time employed (self-employed, employee)
C_Q07	2	Part-time employed (self-employed, employee)
C_Q07	3	Unemployed
C_Q07	4	Pupil, student
C_Q07	5	Apprentice, internship
C_Q07	6	In retirement or early retirement
C_Q07	7	Permanently disabled
C_Q07	8	In compulsory military or community service
C_Q07	9	Fulfilling domestic tasks or looking after children/family
C_Q07	10	Other
C_Q07	96	Valid skip
C_Q07	97	Don't know
C_Q07	98	Refused
C_Q07	99	Not stated or inferred
D_Q12a	1	No formal qualification or below ISCED 1
D_Q12a	2	ISCED 1
D_Q12a	3	ISCED 2
D_Q12a	4	ISCED 3C shorter than 2 years
D_Q12a	5	ISCED 3C 2 years or more
D_Q12a	6	ISCED 3A-B
D_Q12a	7	ISCED 3 (without distinction A-B-C, 2y+)
D_Q12a	8	ISCED 4C
D_Q12a	9	ISCED 4A-B
D_Q12a	10	ISCED 4 (without distinction A-B-C)
D_Q12a	11	ISCED 5B
D_Q12a	12	ISCED 5A, bachelor degree

Variable Name	Value	Value Label
D_Q12a	13	ISCED 5A, master degree
D_Q12a	14	ISCED 6
D_Q12a	15	ISCED 5A bachelor degree, 5A master degree, and 6 (without distinction)
D_Q12a	96	Valid skip
D_Q12a	97	Don't know
D_Q12a	98	Refused
D_Q12a	99	Not stated or inferred
I_Q08	1	Excellent
I_Q08	2	Very good
I_Q08	3	Good
I_Q08	4	Fair
I_Q08	5	Poor
I_Q08	6	Valid skip
I_Q08	7	Don't know
I_Q08	8	Refused
I_Q08	9	Not stated or inferred
ISIC1C	A	Agriculture, forestry and fishing
ISIC1C	B	Mining and quarrying
ISIC1C	C	Manufacturing
ISIC1C	D	Electricity, gas, steam and air conditioning supply
ISIC1C	E	Water supply; sewerage, waste management and remediation activities
ISIC1C	F	Construction
ISIC1C	G	Wholesale and retail trade; repair of motor vehicles and motorcycles
ISIC1C	H	Transportation and storage
ISIC1C	I	Accommodation and food service activities
ISIC1C	J	Information and communication
ISIC1C	K	Financial and insurance activities
ISIC1C	L	Real estate activities
ISIC1C	M	Professional, scientific and technical activities
ISIC1C	N	Administrative and support service activities
ISIC1C	O	Public administration and defence; compulsory social security
ISIC1C	P	Education
ISIC1C	Q	Human health and social work activities
ISIC1C	R	Arts, entertainment and recreation
ISIC1C	S	Other service activities

Variable Name	Value	Value Label
ISIC1C	T	Activities of households as employers; undifferentiated goods- and services-producing activ. of households for own use
ISIC1C	U	Activities of extraterritorial organizations and bodies
ISIC1C	9995	No paid work for past 5 years
ISIC1C	9996	Valid skip
ISIC1C	9997	Don't know
ISIC1C	9998	Refused
ISIC1C	9999	Not stated or inferred
ISCO1C	0	Armed forces
ISCO1C	1	Legislators, senior officials and managers
ISCO1C	2	Professionals
ISCO1C	3	Technicians and associate professionals
ISCO1C	4	Clerks
ISCO1C	5	Service workers and shop and market sales workers
ISCO1C	6	Skilled agricultural and fishery workers
ISCO1C	7	Craft and related trades workers
ISCO1C	8	Plant and machine operators and assemblers
ISCO1C	9	Elementary occupations
ISCO1C	9995	No paid work for past 5 years
ISCO1C	9996	Valid skip
ISCO1C	9997	Don't know
ISCO1C	9998	Refused
ISCO1C	9999	Not stated or inferred
IMGEN	1	1st generation immigrants
IMGEN	2	2nd generation immigrants
IMGEN	3	Non 1st or 2nd generation immigrants
IMGEN	4	Non-immigrant and one foreign-born parent
IMGEN	9	Not stated or inferred

C

CORRELATION MATRIX

	Female	Experience	isco1c	RequiredYears	Wage	OEDU_YRS	UEDU_YRS	OEDU_DUMMY	UEDU_DUMMY
Female	1	-0.122	-0.050	-0.023	0.017	-0.001	-0.071	0.004	-0.055
Experience	-0.122	1	-0.126	0.184	0.022	-0.176	0.188	-0.154	0.196
isco1c	-0.050	-0.126	1	-0.627	0.038	0.282	-0.029	0.219	-0.061
RequiredYears	-0.023	0.184	-0.627	1	-0.039	-0.628	0.199	-0.485	0.234
Wage	0.017	0.022	0.038	-0.039	1	-0.008	-0.009	-0.009	-0.011
OEDU_YRS	-0.001	-0.176	0.282	-0.628	-0.008	1	-0.225	0.806	-0.267
UEDU_YRS	-0.071	0.188	-0.029	0.199	-0.009	-0.225	1	-0.280	0.844
OEDU_DUMMY	0.004	-0.154	0.219	-0.485	-0.009	0.806	-0.280	1	-0.331
UEDU_DUMMY	-0.055	0.196	-0.061	0.234	-0.011	-0.267	0.844	-0.331	1