08-04-2018

The effects of over-education among Pedagogical and Educational Science graduates

With three labour market theories: Human Capital Theory, Job Competition Model and Assignment Model

Postma, L.J. (Lars)
RADBOUD UNIVERSITEIT NIJMEGEN

SUPERVISED BY DR. A. DE BEUCKELAER
SECOND READER: DR. J.J.L.E. BÜCKER

STUDENT NUMBER: S4082338
ADDRESS: 6521 DC 103 NIJMEGEN
PHONE: 06-22913227
E-MAIL: LJ.POSTMA@GMAIL.COM

Table of Contents

Foreword

This master thesis is written between January 2017 and April 2018 and supervised by dr. A. de Beuckelaer. The difficulties for my girlfriend to find a job on her educational level, has motivated me to write my thesis on the subject of over-education. This thesis is realized in order to complete the final stages of my Business Administration master at Radboud University Nijmegen.

I want to thank a number of people that made it possible for me to finish my master thesis successfully. First of all I want to give special thanks to my girlfriend Anna Jonker, a huge support in general and without her in my life it would be a whole lot more difficult. Secondly I want to thank my supervisor dr. A. de Beuckelaer. The supervising of this thesis did not always went smooth, but I can only be proud when I see the final result of the thesis. With his help and notes, I improved my skills and increased the quality of this thesis. At last, I want to thank all the persons who have spent their time to make it possible for me to graduate: dr. J.J.L.E Bücker who agreed to be my second reader, Lieke Gehling & Kirsten Moonen for testing and critically reviewing the questionnaire and a special thanks to all of the respondents that filled in the questionnaire.

Summary

In this study, we investigate how over-education among DPES graduates sits within the context of three labour market theories: Human Capital Theory (HCT) Job Competition Model (JCM) & Assignment Model (AM). The difference in interpretation among the three labour market theories causes different testable hypotheses with respect to wage and individual differences. This study tests these testable hypotheses on over-education among Dutch Pedagogical and Educational Science (DPES) graduates. This research focused on DPES graduates due to the bad labour market conditions wherein there are too many graduates in comparison with the amounts of jobs available. Over-education among the DPES graduates is widely represented (38.1% in our data) excluding the unemployed DPES graduates (6.7% in our data). The hypotheses on wage are tested with the ORU-specification with the dummy model in a linear regression. The results on wage show that over-educated DPES graduates earn significantly less as correct educated DPES graduates. These results confirm the hypotheses from the JCM and the AM, but are contradicting the HCT. The second hypotheses are about individual differences and if over-education is caused by different individual's capabilities. Useful individual capabilities for DPES graduates are derived from O'NET and thereby this is one of the first studies that included explicit measures for individual capabilities. The total amount of useful capabilities is reduced to six factors with a factor analysis to test these capabilities in a logistic regression. The results of the logistic regression show that over-education is not caused by a lack of certain capabilities and thereby the results are contradicting the AM but are in line with HCT and JCM. Over-educated individuals score even significantly higher on one of the capabilities, namely Analytical Thinking. Our results show that over-educated DPES graduates earn significantly less but are not worse in any of the useful capabilities, favouring the JCM interpretation of the labour market. On balance, the evidence from our questionnaire among Dutch Pedagogical Educational Science graduates tends to support the *Job Competition* interpretation of the labour market, implying that the job characteristics for determining wage are important. Previous research (McGuinness, 2006; Battu et al., 1999; Hartog, 2000) tends to favour the *Assignment* interpretation of the labour market, wherein both individual and job characteristics are important. This research incorporated explicit measures of beneficial capabilities for DPES graduates, however none of the capabilities decreased the chance of being over-educated. Our results highlight the importance of a good fit between the organization and the over-educated worker. Over-education leads to lower job satisfaction with the result of lower productivity, more absenteeism and scarring. Organisations should

enhance the opportunities for over-educated workers to improve their capabilities to increase the job satisfaction and to protect from scarring the individual. Recruitment has an important role in determining the fit between the applicant and the organisation, wherein there are possibilities for organisations to find a different position (or even create a new one) for the OE worker within the company, to take advantage of his or her capabilities.

1. Introduction

The last four decades have witnessed rapid and massive changes in educational participation. In particular an increase in participation in higher education and dramatic increases in the number of men and women at work holding university qualifications (Dolton & Silles, 2006). According to Hartog (2000), this strong expansion of participation in education has outpaced the increased demanded level of education. Or in other words, the amount of educated workers is increasing faster than the amount of jobs available to them. The situation of an abundance of educated workers leads to a bad labour market, wherein it becomes more difficult for educated workers to find a suitable job for their capabilities. The bad labour market causes people to stay longer in unemployment due to difficulties finding a job, but also causes people to accept a job below their capabilities, i.e. over-educated (OE) workers. Accepting a job below one's capabilities can be used as a stepping stone to get a more suitable job thereafter, but there are various consequences of being OE.

Over-education is investigated in various fields of studies. Business administration's interest is on individual and organizational outcomes, while economists investigate the underutilization of the labor force and its effects on the labor market and wages. (Mckee-Ryan & Harvey, 2011). A clear pattern of negative relationships emerges with over-education (both economically as personally) from various studies. OE workers are found to be less satisfied with their job than CE workers (Quintini, 2011), higher chance of depression (Dooley, Prausse & Ham-rowbottom, 2000), lower organizational fit (Maynard et al., 2006), lower tenure (Jones et al., 2008), lower wage (McGuinness, 2006) and in general lower job performance although researches failed to reveal a clear pattern (Mckee-Ryan, Harvey, 2011). In this study an interdisciplinary view is used by testing the assumptions with respect to the economic effect of over-education on wage as well as the individual effect to whom over-educations occurs. Which allows us to make practical implication suggestions for organizations that can be implemented due to over-education.

The presence of over-education is widespread, especially in the U.S. and Europe (Tsai, 2010). The meta-analysis of country-studies by Quintini (2011) reveals that the amount of OE workers varies between 10% and 35% in most OECD countries, but there are significant variations across countries and socio-demographic groups. Most research about over-education in Europe is conducted with data from United Kingdom, Germany, Netherlands and Spain since over-education seems to be most present in those countries. In the Netherlands (as most of those countries), an explanatory reason for the presence of over-education is the increased amount of participation in education. The amount of Dutch males that left school with only six years of basic education has reduced from 10% in 1961 to less than 1% in 1981 (Hartog & Oosterbeek, 1988). In addition, the amount that finished school with a bachelor or master degree keeps increasing compared to older generations (Table 1).

Table 1
Amount of the Dutch population that has a higher degree in 2013

	Age				
	25-35	35-45	45 – 55	55 - 65	All
Bachelor degree	19,3%	16,8%	16,3%	13,5%	14,6%
University degree	10,7%	9,9%	8,4%	6,5%	7,6%

Data retrieved from CBS, 2017

The increased amount of high-educated graduates has led to current high-educated graduates being more vulnerable for over-education (Battu et al., 1999). However the chance of being OE is not equal for every graduate and varies on the field of study. The results of Barone & Ortiz (2011) reveal that graduates from humanistic fields have a significantly higher chance of being OE. In the Netherlands, one of the graduates groups in the humanistic fields that seems to suffer from over-education is Dutch Pedagogical and Educational Science (DPES) graduates. A relative high amount of DPES graduates in combination with a decrease in the amount of suitable workplaces results in limited opportunities for DPES graduates (Julen, 2016).

Till now there is no consensus about the explanation for the existence of over-education and there are various theories that try to explain the discrepancy between someone's actual educational level and the correct educational level of the job (Chiswick & Miller, 2009). In this paper three of the key labour market theories[1] that explain over-education are used: human capital theory, job competition model and assignment model. Each of these theories has a different interpretation of the labour market and will be briefly explained below.

The Human Capital Theory (HCT) uses a mathematical interpretation that rests on the assumption that wage of the individual is equal to the productivity of the individual. The total amount of knowledge and skills an individual has gathered by education and experience is called the human capital and determines the productivity. Education increases the human capital of the individual and is seen as an investment in the individual itself, which leads to higher future wages for the individual. The future increase in wages compensates and rewards

[1] Although there is a difference between model and theory, for simplicity we reference the HCT, JCM and AM as three theories.

the initial education costs made, like money and time. Thereby the HCT alleges that a more educated population is a more productive population (Olaniyan & Okemakinde, 2008). According to the HCT, the wage an individual receives is embedded in the human capital of the individual and therefore over-education is the result of 1) a deliberate choice because the lower level job is a good investment opportunity or 2) omitted human capital components (Hartog, 2000).

The Job Competition Model (JCM) uses a competitive interpretation where individuals compete with their qualifications for the allocation of the jobs. Individuals compete for job opportunities based on individual's amount of training costs and the individual with the most amount of training costs will get the job, i.e. the position in the hiring que is determined by the amount of training costs (Quintini, 2011). Participating in an education leads to an increase in training costs and thereby participating in education leads to an upwards shift in the queue. An increase in training costs does help with hiring, but it does not affect the productivity or wage of the job. The observation behind the model is that a majority of workplace skills are acquired through on-the-job training and determined by the job characteristics, regardless of the attributes of the individual who fills it (McGuinness, 2006). With respect to wage, the firm determines what the value of the job is worth to them and the value is determined by job characteristics (Dolton & Vignoles, 2000). Since the productivity is tied into the job, OE individuals are equally productive as workers that perform a similar job and thereby receive the same wage. CE individuals (with a similar amount of education as OE individuals), perform a more demanding job and therefore will receive a higher wage.

The Assignment model (AM) uses an assignment interpretation that incorporates both elements of the capabilities of the individual and the job characteristics. In this model the productivity of the individual is positively related to education, yet not all similarly educated workers are equally productive in all jobs (Linslay, 2005). Individuals have a comparative advantage in jobs where the capabilities of the individual are an accurate match with the job characteristics. The model is based on the assumption that there is an allocation problem with heterogeneous workers that need to be assigned to heterogeneous jobs, which leads to inaccurate matches between the capabilities of the individual and the job characteristics (Büchel & Pollmann-Schult, 2004). Over-education arises when individuals choose a field of study or job which does not conform to their personal capabilities (Linslay, 2005). The level of skill an individual has in a job depends on the amount of education, but also on the match between individual's capabilities and those needed for the job. In the AM the most skilled individual is assigned to the most complex job and this process will continue until the lowest skilled individual is assigned to the least complex job. The wage is positively related to the complexity of the job. If individual's capabilities are not in line with those needed for the job, a less complex job is distributed and over-education occurs (Linsley, 2005). OE individuals perform a less complex job and receive a lower wage.

The difference in interpretation among the three labour market theories causes different testable hypotheses with respect to wage and individual differences. This study will test these testable hypotheses on over-education among Dutch Pedagogical and Educational Science (DPES) graduates. The hypotheses about wage are answered with a linear regression, wherein the wage for OE-DPES graduates is compared with the wage for CE DPES graduates with a dummy variable. The hypotheses about individual differences are tested with a logistic regression, wherein the individual capabilities of DPES graduates are compared with the individual capabilities of CE-DPES graduates. This thesis will focus on how over-education among DPES graduates sits within the context of the three labour market theories (HCT, JCM and AM).

Several studies have attempted to compare different labour market theories with respect to over-education. The research most related to this thesis is the work of McGuinness (2006) who investigated the wage on North-Ireland graduates and assessed the consistency of the HCT, JCM & AM. The results of McGuinness (2006) tend to support the AM, implying that both the individual and the job characteristics are important for determining wage rates in the economy. Comparable work is done by Sloane et al. (1999) who similarly tested the consistency of the HCT, JCM & AM among a British labour market dataset and concluded that both the HCT and JCM are rejected in favour of encompassing the AM. Hartog (2000) also performed an analysis in which he compared various labour market models including the HCT & AM in different countries including the Netherlands. The results of Hartog (2000) indicate that the AM corresponds the best, however he emphasizes that more empirical research is needed. The empirical results tend to favour the AM, however there are concerns that the absence of controls for individual capabilities overstate the effect of over-education and understate the role of human capital (McGuinness, 2006). In this study explicit useful individual capabilities are identified for DPES graduates and therefore the effect of individual capabilities on over-education can be taken into account.

In the present study, we investigate how over-education among DPES graduates sits within the context of the three labour market theories (HCT, JCM & AM). Each labour market theory is profusely explained in the next section (section 2.2), where each section is finished with the corresponding hypotheses about the effects of over-education on wage and individual differences. Subsequently in the method section is explained how the hypotheses are tested (section 3), but also how over-education is measured (section 3.4.1) and the measures of useful individual capabilities (section 3.4.2). In the fourth section the results of the analysis are presented, whereby first the data is examined (section 4.1) and the amount of capabilities is reduced into underlying dimensions with a factor analysis (section 4.2). Thereafter the hypotheses for wage are tested with a linear regression (section 4.3) and the hypotheses for individual capabilities are tested with a logistic regression (section 4.4). In the fifth section the results are discussed and recommendations are made for further research after which this study is concluded with the conclusion (section 6).

2. Literary overview of Over-Education

In the literature there is no accepted unified theory of over-education. There are a number of different labour market theories that try to explain over-education. A significant portion of the literature on over-education considers how over-education sits within the context of these labour market theories (McGuinness, 2006). Each of these labour market theories has a different interpretation of the labour market which offers certain insights and partial explanations for over-education, which in turn generate a number of testable predictions (Sloane et al. 1999). To come to these testable predictions, firstly three labour market theories are explained: Human Capital Theory (HCT), Job Competition Model (JCM) and the Assignment Model (AM). This thesis will focus on how over-education among Dutch Pedagogical Education Science sits within the context of these three labour market theories. To get to there, over-education will be clarified and defined followed by the individual explanation of the three market theories (HCT, JCM & AM)

2.1 Definition of Over-Education

The literature on the subject over-education is widespread and crosses not only different studies (psychology, sociology, management & economics) but also different terminology: ‘underemployment’, ‘skill surplus’, ‘over-education’, ‘over-qualification’ and ‘underutilization’. The difference between the different terminologies is subtle, but the commonality is that individuals have the feeling that they are capable of performing a more demanding job and working at their educational level seems also desirable, at least for the individual itself. Whether they have the right qualifications for a job, gained the appropriate skills by being trained and educated or gained enough experience, over-education is about the mismatch between their skills and the job's requirements.

To speak of a mismatch between the skills of the individual and the skill requirement for the job, both individual skill level and the requirements of the job need to be determined. The advantage of a certification from an education is an objective measurement which guarantees a minimum level of individual's skills, individuals in possess of a certificate transcends at least that minimum level. Using the educational level is an accepted and objective measurement of an individual's skill level and will be the used from now on. A simple definition of over-education would be that actual education is greater than required for the job (Hartog, 2000). The most comprehensive approach to define over-education is offered by Feldman (1996), who defined five dimensions that are used to judge whether someone is over-educated. A conformation with any of these five dimensions would qualify a person to be over-educated:

1. More formal education than their current job requires
2. Involuntarily employed in a different field from that in which they received their formal education
3. Passes higher-level skills than the job requires
4. Involuntarily engaged in part-time, temporary or intermittent employment
5. Earning 20 percent less than the average earnings of their graduating cohort in the same major or occupation track (adopted from Scurry & Blenkinsopp, 2011, p. 646)

The advantage of using these dimensions is that they are more comprehensive and also take into account the persons who got crowded out in a different field (dimension 2), the wage (dimension 5) or those who unwillingly have a lower (or higher) amount of workhours than they desire (dimension 4).

Chevalier (2003) adds satisfaction with the match between education and job as another dimension of whether someone is overeducated. The distinction is made between apparent over-educated (AOE) where an individual is in a job below their educational level but is satisfied with the match between their qualifications and job, or genuine over-educated (GOE) where an individual is in a job below their educational level and is dissatisfied with their match. In general, as displayed with the dimensions of Feldman (1996), over-education is linked to negative effects: low usage of individual's capabilities (dimensions 1 & 3), work in a different field than desired (dimension 2), undesired amount of workhours (dimension 4) and lower wage (dimension 5). Being OE as individual seems to be undesirable, however some individuals purposely choose a job below their educational level and are satisfied with this situation. One of the explanations for AOE is that individuals look for jobs in the regional market rather than the global one. Being restricted to a region limits the choice opportunities of a job for the individual. A large portion of research of this effect is done with married women where the woman is limited to the regional market of the spouse. The results of over-education among those women are mixed (Büchel & van Ham, 2003), but when someone willingly accepts a job below their educational level and is satisfied with this match there is no need to label them as GOE. Combining the dimensions of Feldman (1996) & Chevalier (2003) leads to the definition that will be used: *An individual is qualified as overeducated when they have more formal education than their current job requires, or is employed in a different field than their study, and is dissatisfied with this match.* This definition will be used for the measurement for over-education on DPES graduates (section 3.4.1). As the definition is clear, the next step is to explain the different labour market theories. As mentioned in the introduction, there is no consensus about the existence of over-education. The Human Capital Theory uses a mathematical interpretation, the Job Competition model uses a competitive interpretation and the Assignment model uses an assignment interpretation of the labour market. Since the three theories have different interpretations of the labour-market, the explanations for the existence of over-education and the consequences of over-education on wage differ. Below is each theory separately explained and each theory will be finished with the corresponding hypotheses.

2.2 Human Capital Theory

The HCT is a mathematical framework, characteristic of the neo-classical view, which illustrates the relationship between education and wage (Mincer, 1974). To have a clear picture of the HCT it is helpful to understand one of the basic assumptions of the neo-classicalists. Neo-classicalists assume that all individuals seek to maximize their own interest and all the roots of social phenomena could be found in individuals' behaviours (Tan, 2014). The basic idea of the HCT is that wage is determined by the individual productivity of an employee. A firm pays the individual exactly what the individual is worth to them, i.e. individual's productivity. Since the aim of a firm should be cost minimization and profit maximization the employee receives no more than their productivity, but also not less since another firm would snatch these employees. To satisfy this condition the assumption of perfect capital markets needs to be made (Mincer, 1974). Perfect capital markets is a state wherein all firms have full information about each individual's productivity and behave perfectly rational in the wages the firms offer. If perfect capital markets are assumed then the wage of an individual is equal to individual's productivity. The level of the productivity is fully embodied in the human capital of the individual and does not depend on the requirements of the job according to the HCT. The human capital is the total sum of the knowledge, experience and intelligence of the individual (Groot, 1996).

Since individuals want to maximize their own interest (assumed by neo-classicalists), increasing human capital is expedient due to the increase in productivity (and then wage). Increasing human capital through education or other trainings however is not without costs. Direct costs as institution's fee need to be paid, but also indirect costs as time. The actual lifetime earnings (Y) at time t depends not only on the earnings at time t, but also on the investment costs (I) at time t. Individuals try to find the optimal balance between the wage received (wK) at time t and the costs that are made to increase productivity (I) at time t.

Equation 1: The total life-time earnings of an individual

$$Y(t) = wK(t) - I(t)$$

The optimal balance is at the point where the individual maximizes present value of his lifetime earnings ($Y_{(t)}$), by appropriately choosing the time they spend on education and work. If $K_{(t)}$ is each individual's human capital at time t and w is the wage per unit of capital then multiplying both terms results in individual earnings at time t ($wK_{(t)}$). Note that the productivity of an individual is based on individual's level of human capital which results in an individual wage ($wK_{(t)}$). Since Neo-classicalist's assume that individuals behave rational and maximize their own interest, individuals seek to maximize the present value of lifetime earnings. When the present value of lifetime earnings are maximized then equally the lifetime earnings of the individual are maximized. To maximize Y, the amount the individual spends on work and the amount that the individual invests in their individual human capital needs to be chosen appropriately. An investment has a benefit (increased future earnings) but is illustrated as a cost (foregone earnings). Y is maximized when an additional euro spend in education does not lead to more than an additional euro in wage or in other words equating the costs and benefits at margin in each year results in a maximum. The point in which both the margin costs and benefits are at margin in each year is the optimal strategy to maximize the present value of earnings (Polacheck p. 175, 2008). Maximizing the present value of lifetime earnings leads to the following formula:

Equation 2. The present value of lifetime earnings of an individual (maximized)

$$PV = Max \int_0^T Y(t)e^{-rT} dt$$

Mincer (1974) was the first to derive an empirical formulation of an individual's earnings over the lifecycle at any time (Equation 3). The formula, also known as the Mincer Earnings Function, can be depicted as a concave function of one's labour market experience. The earnings at any time depend on the schooling (S) in years each individual has, the labour market experience (t), the initial earnings capacity (a_0) and a residual term (Polacheck, 2008). A simplification of the model is that an individual only has the choice between education and work. Dividing the time with part-time work or studying part-time is not encountered in the model, as a delay in a study or leisure.

Equation 3: Mincer Earnings Function

$$\ln Y_1(t) = a_0 + a_1 S_i + a_2 t_i + a_3 t_i^2 + \varepsilon_i$$

One of the achievements of Mincer is that he found an equilibrium position where identical individuals optimally invest in human capital to maximize the present value of their future earnings (Lemieux, 2006). An individual will keep studying until another year of education has no increased benefits to the present value of his total earnings. The formula of individuals earning at any time (Equation 3) is mainly focused on the amount of schooling in years an individual has and the amount of work experience in years the individual possesses. Completing an education leads to an increased wage, since the productivity of the individual increases. Earning less than another individual is the result, according to the Mincer Earnings Function, of a lower amount of human capital. The model applies equally to all persons and assumes homogeneity of individuals. Individuals that participated a similar amount of years in education have an identical amount of increase in human capital (and wage) compared to each other. Differences between individuals' wage that participated a similar amount of years in education is caused by omitted human capital.

Hypotheses Human capital theory

In the previous section there is determined that according to the HCT the wage is completely embedded in the productivity of the individual and the productivity is embedded in the human capital of the individual (the knowledge, experience and intelligence of the individual). The amount of human capital is not static, but could be increased by gaining knowledge through education or experience in the job. With recent graduates job experience is barely gained and the amount of human capital depends largely on the amount of years an individual spends on education. This implies that graduates that follow the same full-time educational program will have an identical increase in human capital and thereby the same increase in wage. DPES graduates will have largely the same amount of human capital, besides the small difference in intelligence. Due to the largely similar amount of human capital DPES graduates will be almost equally productive and thereby earn an identical amount of wage. Over-education does therefore have no influence on the wage, because as mentioned at the start of the section the productivity is embedded in the human capital. The same concept applies to under-educated (UE) individuals. The amount of human capital determines the productivity and thus the wage. The hypothesis by HCT for over-education on wage is as follows:

Hypothesis 1a: According to the Human Capital Theory, the wage for OE individuals is equal in comparison with individuals that are CE.

Proponents of the HCT state that the amount of increase in human capital is equal to all individuals. Hence, there is a homogeneous effect on individuals' productivity, if individuals make the same investment on human capital. Over-education is not caused by individual capabilities. Over-education according to the HCT is the result of 1) a deliberate choice because the lower level job is a good investment opportunity, or 2) omitted human capital components (Hartog, 2000).. The hypothesis for the explanation of over-education is as follows:

Hypothesis 1b: According to the Human Capital theory, OE is not caused by individuals capabilities (homogeneity)

2.3 Job Competition Model

The Human Capital Theory presents a valid suggestion why someone would invest in education. An education leads to higher future earnings, which makes up for the initial sacrifice. But when these future earnings are lacking, the incentive to invest in education disappears. In this respect over-education can be seen as a puzzle, why would someone invest in education when they cannot fully utilize (Brynin & Longhi, 2009). The Job Competition Model (JCM) has a different view on over-education. The assumption behind the model is that there is a skill level attached to a job, regardless of the attributes of the person who fills it. Whilst it seems reasonable, it is contradicting with the HCT that assumes that an individual produces a productivity, irrespective of the job in which that individual works (Green &

McIntosch, 2007). The JCT thereby focuses on the demand side (the job itself) and assumes that the productivity and wage are tied into the jobs, instead of the individual (Muysken & Ter Weel, 1999). The wage an individual receives depends therefore not on the productivity or education an individual has but is tied in the job.

The JCT states that the amount of training or experience an individual possesses is not definable in the productivity, however it is definable when being hired. Individuals are ranked by the amount of training each individual has in a so-called labour queue. The most skilled individual (with the most training costs) are in front of the labour queue and the least skilled individual (with the lowest amount of training cost) is at the end of the labour queue. The available individual with the most training costs will be hired, i.e. the relative education level determines individual's employment chance (Muysken & Ter Weel, 1999). Completing an education leads to an upwards shift in the queue and increases the employment chance. The increased employment chance would also be the reason why someone would invest in education, but it gives no guarantees on wage or the type of job. It is possible that with a limited amount of high-skilled jobs and an abundance of high-skilled workers the remaining high skilled workers still have to accept a job below their capabilities. The same concept applies vice versa, with an abundance of high-skilled jobs the lower skilled workers have to fill in the spots of the high-skilled jobs. Over-education can arise by an abundance of high-skilled workers with a limited amount of jobs available to them. This explanation of over-education is in line with Hartog (2000) who argued that due to increased participation in education the amount of skilled workers increased faster than the demand of skilled workers and over-education arises. The prediction of the JCT is (in contrast to the HCM) that being over-educated does not lead to a surplus in earnings. The job characteristics determine earnings and qualifications are only important for the allocation of the jobs (Quintini, 2011). Investing in education will protect your place in the queue, but still gives no guarantees for a job wherein individuals utilize all their capabilities. Hence there always exists qualification mismatches, except when the labour market and the job market are in perfect equilibrium. The existence of over-education wherein people have more skills than the job demands is understandable with the JCT, however there is a difference between a minor mismatch and a major mismatch. A major mismatch according to the HCT does matter less, since the wage is determined by individual's productivity. With the JCT the wage is determined by the job characteristics and not someone's individual capabilities, having a mismatch does not lead in any way to an increased wage.

Hypotheses Job Competition Model

As written in the previous section, with JCM the productivity is complete embedded in the job characteristics and not determined by the capabilities of the individual. An OE individual will not be more productive than an individual who practice the same job but obtained less education. Since both individuals will be equally productive there is no difference in wage. The wage depends on the complexity of the job (determined by the educational level), where the most complex job will pay the most wage. OE individuals work at a less complex job as CE individuals and thereby earn a lower wage. UE individuals will work at a more complex job as CE individuals and thereby earn a higher wage. The hypothesis for wage for over-education is as follows:

Hypothesis 2a: According to the Job Competition Model, the wage for OE individuals is lower in comparison with individuals that are CE.

The wage an individual receive depends on the complexity of the job. A more complex job can be obtained by increasing the amount of training costs which improves the relative position in the queue. The expectation is that recent DPES graduates in general have not obtained any additional training costs except the education they have completed. The amount of training costs for recent DPES graduates are approximately identical. Over-education occurs if there is an abundance of high-skilled individuals and limited amount of high-complex jobs. The determination of who will be hired for the high-complex job is the amount of training costs. Among graduates the amount of training costs are approximately identical, thereby they have a close to identical chance of being hired. Over-education is caused by the amount of training costs, not individual capabilities. The hypothesis for the explanation of over-education is as follows:

Hypothesis 2b: According to the Job Competition Model, OE is not caused by individuals capabilities. (homogeneity)

2.4 Assignment model

The last model discussed in this paper is the Assignment model (AM), a model in the middle ground between HCT and JCT. As discussed above, the HCT assumes that the productivity is completely embedded in the human capital of the individual. The JCT assumes that the productivity is completely embedded in the job and individual’s educational level only matters for job allocation. The AM incorporates elements of both the characteristics of the job and the capabilities of the individual. The model is based on the assumption that there is an allocation problem with heterogeneous workers that needs to be allocated to heterogeneous jobs, which may lead to unemployment or mismatches such as over-education (Büchel & Pollmann-Schult, 2004). Individuals are ranked top-down by the level of quality (skill) they possess and jobs are ranked top-down by quality (complexity). The allocation of workers is distributed according to the ‘zipper-allocation’. The best individual will gain the most complex job and the ‘zipper’ slides down to match both sides, until at least one side is exhausted (Hartog, 2000). The wage distribution is consistent with the complexity of the job, the most complex job will receive the highest wage and the least complex job the lowest wage.

An individual’s position on the zipper is based on the skill level the individual possess. As well as the other two theories, the AM assumes that education is positively related to your skill level. But in contrast to the HCT, individuals with the same level of human capital are not equally productive, the productivity of the individual depends on the job they are matched with. Certain capabilities of the individual can be desired in the job, creating a competitive advantage for that specific individual in that particular job. The result is that individuals with the same completed education are not necessary evenly productive in every job. Each individual has their own capabilities, but different sets of capabilities are favourable in different jobs. The difference between the capabilities possessed and demanded creates the allocation problem with heterogeneous individuals and heterogenous jobs. An individual’s productivity in a job depends thus on the individual’s amount of education, job characteristics and how individuals’ capabilities connect with the job characteristics. If individual’s capabilities are in line with the capabilities beneficial in the job then those individuals will have a comparative advantage in those jobs. Over-education arises when individuals are not aligned in the jobs which those individuals have a comparative advantage (Linsley, 2005). OE individuals are assigned to a job with a lower amount of complexity causing difficulties for individuals to fully utilize their skills. OE individuals impose a ceiling on productivity (Quintini, 2011). OE individuals are however more productive than their co-workers with the same job due to the higher educational level. The difference with JCM is productivity is not only embedded in the job but also in the individual. Additional education will be beneficial to increase productivity in the current job, but the capabilities of the current job will not have the potential to completely utilize the additional education. The productivity of an individual in a job is determined by both the capabilities of the individual and the complexity of the job according to the assignment model.

Hypotheses Assignment model

In the theoretical section of the AM (2.3) the 'zipper-allocation' is explained, whereby the most skilled worker will have the most complex job and thereby receiving the highest wage. There will be 'zipped' down until either the jobs or the workers are exhausted resulting in the lowest skilled worker receiving the least complex job with the lowest wage. The relative place on the zipper is positively related to the wage an individual receive. OE individuals possess a less complex job than CE individuals and thus the lower complexity for OE individuals causes a lower wage. The same story applies to UE but then vice versa. UE workers will possess a more complex job than CE individuals and thus the higher complexity for UE individuals causes a higher wage. The hypothesis about the wage for over-education is as follows:

Hypothesis 3a: According to the Assignment model, the wage for OE individuals is lower in comparison with individuals that are CE.

The wage individuals receive depends on the complexity of the job. The AM states that individuals' capabilities are beneficial to receive a more complex job and create a comparative advantage in their job field. An individual is more productive when individual's capabilities are in line with the capabilities beneficial for the job. The AM hypothesizes that there is heterogeneity among the capabilities of the individuals and over-education occurs when individual's capabilities are less beneficial in the job field (Linsley, 2005). OE individuals have chosen a job direction that not aligns with their capabilities. The amount of training costs (named in JCM) or human capital (named in HCT) has a positive influence on gaining a more complex job, but additional to those theories individual capabilities matters as well.

Hypothesis 3b: According to the Assignment model, OE is caused by individuals capabilities (heterogeneity).

2.5 Overview Labour market theories

The primary difference between the three explanations of the labour market for over-education is with respect to wage and individual differences. In table 2 a clear picture and good overview is presented of the different theories. All of the three theories have their own explanation for over-education, however there is communality between the hypotheses. The hypotheses from JCM and AM on wage match and the hypotheses from HCT and JCM on differences match. Both hypotheses on wage and individual differences need to be tested to see how over-education sits within the context of these three labour market theories. This study focuses on the differences between over-education and correct education, however UE is added as well. If the wages are equal for over-education and correct education, in combination with over-education not being caused by individual differences, the HCT is preferred. If the wages are lower for over-education in comparison with correct education, in combination with over-education not being caused by individual differences, the JCM is preferred. If the wages are lower for over-education in comparison with correct education, in combination with over-education being caused by individual differences, the AM is preferred. The differences in wage has been studied multiple times and an overview is made by Quintini (2011), however individual differences are often assumed instead of investigated (Sloane, 2003). The contribution to the literature lies primarily with testing the hypotheses on differences, but the hypotheses on wage are also tested.

Table 2
Overview of hypotheses for HCT, JCM & AM

	Wage	Individual differences
Human Capital Theory	OE = CE = UE (H1a)	No effect (H1b)
Job Competition Model	OE < CE > UE (H2a)	No effect (H2b)
Assignment Model	OE < CE > UE (H3a)	Cause OE (H3b)

3. Method

In the method section the used research design is presented to empirically test the hypotheses on wage for over-education (H1a, H2a & H3a) and individual differences (H1b, H2b, H3b). To test the hypotheses an electronic questionnaire is used. The measurement of both the cause for over-education and the effects on wage of over-education involves a number of issues that are crucial in assessing the true significance of the phenomenon over-education. These issues are with the different measurement methods of over-education and the lack of sufficient controls within the wage equation framework (McGuinness, 2006). The method section starts with the researched target group and procedure of the questionnaire (section 3.1). The equations to test the hypotheses are presented next: the wage equation framework for the hypotheses on wage (section 3.2.1) and the heterogeneity equation for the hypotheses about individual differences (section 3.2.2). In section 3.4 the issues with the different measurements methods for over-education are explained and the actual measurement of over-education as the other variables are displayed in section 3.5. The used questionnaire can be found in Appendix (section 7.1).

3.1 Researched target group and Procedure

Data for this research is collected from Dutch Pedagogues and Educational Science (DPES) graduates from Radboud University through a quantitative electronic survey done with Qualtrics and distributed on the social network LinkedIn. DPES graduates are educated towards becoming an educationalist, a specific occupation. The Dutch Society of Educationalists (NVO) states that the key tasks of educationalist are: pedagogical research, diagnostics, treatment and counselling. The work fields for educationalists are mainly: education, youth care, disability care, mental healthcare, policy and research. The specific knowledge of DPES graduates leads to a relative homogenous set of occupations, the knowledge is mainly useable and valuable in the occupation as educationalist. The current labour market for DPES graduates is overcrowded with jobseekers, creating strong competition and low amount of opportunities for graduates. The lack of opportunities leads to graduates accepting a job below their educational level to use that job as a stepping stone (Julen, 2016). The reasons of choosing DPES graduates is the relatively homogeneous set of occupations in which DPES graduates work and the amply presence of over-education among DPES graduates. The relative homogeneous set of occupations has the advantage that when the useful capabilities need to be determined a single set of individual capabilities can be drafted.

To reach as many DPES graduates as possible, the electronic survey has been distributed on the LinkedIn-page for DPES graduates from Radboud University. The distributed survey can be found in the Appendix (section 7.1). LinkedIn is a commonly used business- and employment-oriented social network that can offer possibilities for finding a suitable job. Since LinkedIn positively enhances chances for finding a suitable job, many graduates use this social medium platform. To get a large number of respondents for the survey, graduates were informed about the difficulties on the labour market and the research objectives. In the survey respondents were asked about their career and individual capabilities. To get accurate answers about the respondents career, in the introduction of the questionnaire the advice was given to respondents to keep their resume nearby. The option was offered to respondents to withdraw from the survey and continue later on. This freedom serves to withhold respondents' answers once they become bored, tired or confused during the survey or when they did not know how to answer yet.

3.2 Measuring the wage returns and heterogeneity

With the researched target group and procedure explicitly presented, the next step is to declare how the hypotheses are tested. The hypotheses on over-education are with respect to wage and individual capabilities. The equation on wage (wage-equation) is presented first, which is used to test the hypotheses on wage. The equation on heterogeneity (heterogeneity-equation) is presented secondly, which is used to test the hypotheses on individual capabilities.

3.2.1. The wage-equation

The wage-equation is used to test the hypotheses on wage between OE-DPES graduates and CE-DPES graduates. Most studies that measure the returns on wage for education are based on the Mincer earnings function (Equation 3), also presented in the previous chapter (section 2.2). The Mincer earnings function is based on the HCT which assumes that an individual will participate in education until the point wherein the costs outweigh the benefits of education (Polacheck, 2008). In the original Mincer earnings function the amount of education and experience is measured with amount of years. S is the amount of education in years and t is the experience in years, the dependent variable $Y_1(\mathrm{t})$ is the wage at time t.

In a variant based on the Mincer earnings function S (the amount of years) is replaced by years of **O**ver-education, years of **R**equired education (correct education in this thesis) and years of **U**nder-education, i.e. ORU-specification (McGuinness, 2006). Over-education is the surplus of the amount of years in education and under-education is the deficit of the amount of years and a residual error is added.

Equation 4: Variations on the wage-equation

Variant 1: Mincer earnings function

$$\ln Y_1(t) = a_0 + a_1 S_i + a_2 t_i + a_3 t_i^2 + \varepsilon_i$$

Variant 2: ORU-specification

$$\ln Y_1(t) = a_0 + a_1 S_i^{OverEduc} + a_2 S_i^{ReqEduc} + a_3 S_i^{UnderEduc} + a_4 t_i + a_5 t_i^2 + \varepsilon_i$$

Variant 3: ORU-specification with dummies

$$\ln Y_1(t) = a_0 + a_1 D^{OverEduc} + a_2 D^{ReqEduc} + a_3 D^{UnderEduc} + a_4 t_i + a_5 t_i^2 + \varepsilon_i$$

The ORU-specification measures the return on wage for education as the Mincer earnings function, but in addition the ORU-specification measures the effects of having a surplus or deficit in years of education on wage. Two different variations of the ORU-specification are used within the studies: amount of years (Equation 4.2) and dummy variables (Equation 4.3) for over-education, correct education and under-education. The ORU-specification with amount of years is effective for determining the effects on wage each year, whether it is a surplus or a shortage in years of education. The ORU-specification with dummy variables is effective for comparing the wage for over-, correct or UE individuals with the same amount of education (Quintini, 2011).

To test the hypotheses on the effect of over-education on wage (H1a, H2a & H3a) a linear regression is performed with Ordinary Least Squares (OLS). A linear regression shows the relationship between the dependent variable and the independent variable, OLS predicts the regression line with the lowest distance between the residuals. The coefficient obtained from the OLS regression line displays how the variable wage acts when there is an increase in the independent variable. Previous studies that use the ORU-specification with amount of years resulted in a positive coefficient for return on wage for years of over-education, but the return on wage for years of over-education is smaller than the return of wage of individuals that are working at the correct educational level (McGuinness, 2006). Studies that use the ORU-specification with dummies such as Dolton & Vignoles (2000) and Chevalier (2003) report that OE individuals earn less in general than CE individuals with the same amount of education. The focus of this thesis is the comparison between over-education and correct education and therefore the ORU-specification with the dummy variable is used.

Thus the first hypotheses on wage (H1a, H2a & H3a) can be tested with a OLS regression, but the hypotheses about individual differences (H1b, H2b & H3b) are still remaining. The purpose of the hypotheses on individual differences is to test whether over-education is random or that differences in individual's capabilities are a factor why an individual is OE. The dependent variable in the hypotheses on individual differences is no longer wage but over-education. Since the dependent variable is dichotomous (the respondent is or isn't overeducated) a different equation is needed and therefore the more suitable logistic regression is used.

3.2.2. Heterogeneity equation

The logistic regression is the appropriate regression analysis to conduct when the dependent variable is dichotomous (Peng, Lee & Ingersoll, 2002). A logit regression estimates the effect of independent variables in odds of being OE. An important difference between the theories is the homogeneity in productivity (HCT, JCM) versus heterogeneity in productivity among workers (AM). According to the AM, over-education stems from heterogeneity in productivity caused by the difference in individual capabilities. To test the hypothesis of difference in individual capabilities a logit model is used (equation 5), copied from the study of Chevalier (2003). In this logistic model, the capabilities of OE individuals can be compared to capabilities of either UE or CE individuals and test the relationship between individual capabilities and the chance of being OE. The odds of the logistic regression tell whether a greater level of individual capabilities leads to a greater chance of being OE. To make the comparison with capabilities the dependent variable over-education is transformed into a dummy variable for over-education (OE) with the score of '1' for the OE individuals and '0' for the rest. The independent variables consists of the vectors of individual's capabilities (IC). Both equations to test the hypotheses are explained, the next step is how the variables in these equations are measured.

Equation 5: Heterogeneity equation

$$OE = p_i + e$$
$$Log\left[\frac{p_i}{1-p_i}\right] = \beta_0 + \beta_1 IC_1 + \beta_2 IC_2 + \cdots$$

3.3 Methods for the measurement of over-education

The measurement of over-education has so far relied on three methods to determine whether someone is overeducated for the job: the normative, statistical and the self-declared approach (Quintini, 2011). To determine whether someone is OE, individual's educational level needs to be compared with the correct amount of education for the job. An excess of education in comparison with the correct education level will qualify an individual as OE. UE is inversed and arises when individuals have a shortage of education in comparison with the correct education level. The normative and the self-declared approach measures both the individual's actual level of education and the correct educational level needed. The statistical method uses a short-cut to measure the correct educational level needed. Each approach has its own caveats (Chevalier, 2003), therefore each method will be individually explained with some potential biases.

The self-declared approach uses individuals' opinions on the match between respondents job and educational level. The opinions can be used to determine the correct education level for respondent's job. Examples from previous studies of questions to determine the correct education level of the job are: "how much formal education is required to get a job like yours?" or "if they were applying today, what qualifications if any would someone need to get the type of job you have now?" (Battu et al. 2000). The answer on these questions can be compared with the actual attained education level of the individual. The match between the correct and attained education level determines whether an individual is over-, under- or correctly educated for the job they possess. The self-declared approach is often criticized with respect to the measurement error, to what extent can individuals correctly address the level of education required for the job. The measurement error can occur when respondents apply different criteria or lack benchmarks to compare themselves when assessing their job's correct educational level (McGuinness, 2006). Second, the qualifications needed for a job can be upgraded overtime (i.e. job inflation), if employers are faced by a more qualified pool of candidates (Chevalier, 2000). The job's correct educational level is not static and can be encountered by 'job inflation', which results in different answers within the same job on the question how much formal education is required for their job.

The normative approach uses an a priori presumed correspondence between education and job titles. The presumed correspondence is used to determine the correct educational level of each job. Equal to the self-declared approach the presumed educational level of the job title can be compared with the actual attained level of education of the individual to determine whether an individual is correctly matched. The presumed education level needed for each job title can be found in various indices. An index for the Netherlands is the ARBI-code developed by the Dutch Department of Social Affairs, used in the study of Hartog & Oosterbeek (1988). By knowing the job title of the respondent the corresponding required educational level can be found with the help of the ARBI-code. The downside of using the normative approach is the assumption that every individual with the same job title performs similar tasks and has identical educational requirements (Quintini, 2011). Despite that the job title is identical, the variation between the tasks performed inside the job title can even lead to different educational requirements within the job title. In addition, focusing on job titles has the downside that a change in required educational level (job inflation as example) or other cohort effects lead to sensitivity in the index and therefore in the normative approach.

Both the normative and the self-declared approach identify directly the required educational level to determine the match between an individual's job and educational level, in contrast to the statistical approach which determines over- or under-education by deviating educational levels. The statistical approach calculates the distribution of education for each occupation and the ones that deviate more than some ad-hoc value (for example one standard deviation) are the ones quantified as either over- or under-educated. The advantage of the statistical method is this method still can be used, even when there exists no direct information on the required education. In addition, the statistical approach is sensitive to the level of aggregation necessary to obtain a reliable distribution for each occupation (Chevalier, 2003). The statistical approach is often used as a short-cut to define over-education when the required education is not available in the data. Identical to the normative approach, the statistical approach assumes that respondents with the same job title perform similar jobs and requires equal amount of education.

Each approach for measuring over-education seems to have its own drawbacks and advantages. The choice which method will be used often depends realistically on the data that is available, where the statistical approach could be used as a last-resort. But the fact remains that differences in measurements between studies could lead to difficulties comparing the results of these studies. Consistency across the approaches is desirable, but the study of Verhaest & Omey (2006) indicates that the incidence of over-education largely depends on the measurement applied and in their data over-education varies from 13.8% in the sample with the statistical approach to 52.4% in the sample with the normative approach. The correspondence between the different approaches was remarkably small and ranges from 4% to 50% and these results question if all of these three approaches measure the same concept. Not only in the study of Verhaest & Omey (2006) did the incidence of over-education vary between the different approaches, also the meta-analysis of Groot & Van den Brink (2000) reveals that the incidence of over-education varies between the methods. Groot & Van den Brink state that the statistical approach has significant lower estimates of over-education which is also in line with the results of Verhaest & Omey (2006). These results of Verhaest & Omey (2006) & Groot & Van den Brink (2000) indicate that the method chosen to measure over-education leads to different percentages of over-education, with low correspondence between the different approaches. However, the consequences of the different approaches on the results of the wage-equation (discussed in paragraph 3.1) seems to be little, with low systematic variation between the return on wage for over-education (Groot & Van den Brink, 2000). The conclusion of Verhaest & Omey (2006) is that none of the approaches clearly outperform the alternatives of all variables of interest. So far there is no consensus on which approach should be used or enough evidence that one of the three approaches should be excluded.

The method chosen depends on the practicality of this research, which method seems the most suitable. The researched group Dutch Pedagogical and Educational Science Graduates is chosen on the amplify presence of over-education among those graduates. The statistical method quantifies an individual as OE when it deviates some ad-hoc value from the mean.

The expected amplify presence of over-education implies that in all probability over-education is more present than some pre-determined ad-hoc value. Using the statistical approach would underestimate the true existence of over-education among DPES graduates. As options for this research remain the normative and self-declared approach to measure over-education. In recent studies mainly the self-declared approach is used (Quintini, 2011) which has the advantage that no up-to-date index is required and it does not rely on job titles. The self-declared approach is chosen for the applicability, but the measurement error has to be taken into account.

3.4 Measures and variables

With the approach presented how to measure over-education (self-declared approach) and the equations to test the hypotheses, the next step is to declare how the constructs used in the wage-equation and heterogeneity-equation are measured and transformed into variables.

3.4.1 Measuring over-education, correct education & under-education.

In section 3.3 the self-declared approach is chosen to measure over-education. The definition used in this thesis for over-education is: "*An individual is OE when they have more formal education than their current job requires, or is employed in a different field than their study, and is dissatisfied with this match*". To measure and quantify an individual as OE the match between an individual and their job, the field of the job and the satisfaction of the match need to be obtained. To measure the match between an individual and their job the self-declared approach is used. A respondent is asked: "*How do you estimate your highest degree with regard to your current job?*", based on the question of Frei & Sousa-Posa (2012). The four answer possibilities were as follows "the highest degree is not sufficient", "the highest degree corresponds to your job", "the highest degree is superior to your job" and "the highest degree does not relate to your job". An individual is quantified as 'UE' when the degree is not sufficient, 'CE' when it corresponds, 'OE' when it is superior and 'no match' when the degree does not relate. The category 'no match' reflects individuals with a very poor match between the degree obtained and job. Since individuals with a poor match cannot use their skills gained through education on their job, they are also quantified as OE. Since over-education is the main variable a second question is asked in the survey to the respondents: "*Was the degree gained a requirement in the job specification for your main employment?"*, also used in the study of Battu (2000) & Chevalier (2003). The second question is used to validate the answers of the first questions. Contradictory answers on these questions means that is impossible to quantify respondent's match between job and education and therefore these respondents will be declared invalid. A respondent can only be quantified as '1' in one of the three dummy variables (OE, CE & UE) and automatically will be '0' in the other remaining dummy variables. The last distinction made is between apparent over-educated and genuine over-educated, introduced by Chevalier (2003). An individual that is satisfied with their match between job and education are only AOE, while those who are dissatisfied are GOE. The following questions are asked: "*How satisfied are you with the match between your job and your education*?" and "*In which degree did your education contribute to getting a job*?". The answers on these questions are grouped to generate a dichotomous variable of 'satisfied' and 'dissatisfied'. A combination of dissatisfied and OE is categorized as GOE, and combination of satisfied and OE is categorized as AOE.

3.4.2. Individual Capabilities

The variable Individual capabilities must be a reflection of the capabilities that are beneficial for DPES graduates. To identify capabilities beneficial for DPES graduates the database of O'NET is used, an online database and tool for career exploration and job analysis. O'NET has multiple descriptors to describe an occupation such as: tasks, work behaviours, abilities, skills, knowledge areas and work context. Since DPES graduates are educated towards an Educationalist, the skills described in O'NET for the occupation Educationalist are used. The descriptive skills of O'NET represent a person's level of proficiency or competence to perform a task (Peterson et al., 2001, p. 464). The occupation Educationalist was not present in O'NET, due to language barriers. Therefore Counselling Psychologist (20 identified skills) & School Psychologist (20 identified skills) are chosen, both branches of educationalist. The large relatedness between those occupations is also reflected in the skills needed for both occupations, 19 of the skills described are similar. There are a total of 21 unique skills needed: Active Listening, Instructing, Service Orientation, Reading Comprehension, Social Perceptiveness, Speaking, Monitoring, Judgement and Decision Making, Learning Strategies, Writing, Active Learning, Critical Thinking, Time Management, Coordination, Persuasion, Science (only Counselling Psychologist), Complex Problem Solving, System Evaluation, Management of Personnel Resources & Systems Analysis (only School Psychologist).

The skills needed for DPES graduates are identified but are not yet quantified. Each skill is transformed in a single statement wherein the respondent is asked to answer on a 5-point Likert-scale from 'Strongly Disagree' to 'Strongly Agree'. With the exception of two skills (Monitoring & Learning Strategies) which have two statements to increase clarity, but the two statements are reduced to a single score. In addition some statements are reversed to reduce the yes bias. An overview of the skills and related statements can be found in Appendix (section 7.2). Respondents receive a score for each skill, determined by the amount the respondent deviates from the mean. A shortage of a skill compared to the mean will have a negative effect on the score of the capability, whereas an excess of a skill compared to the mean has a positive effect on the score of the capability. The score represents individual's capability, relative to the other DPES graduates.

3.4.3. Other variables

In the previous sections the operationalization is explained of both the match between respondents job and education level, and individual capabilities. To use the wage-equations the variables wage and experience are needed. This section explains how the remaining variables are constructed, including the categories for the categorical variables (Table 1).

Wage

The net-wage of the respondents is asked in combination with the employment rate to determine the full-time wage of the respondent. The net-wage is asked per month and in step-sizes of €249,-, starting from €500 till €3000. The employment rate is asked directly in a slider option between 0% till 100%. The net-wage per month will be divided by the employment rate to determine the full-time wage per month.

Experience

The amount of experience of the respondent is divided into two categories: 'total amount of work experience' and 'total amount of relevant work experience'. For the wage-equation the total amount of work-experience is needed. To determine the total amount of work-experience the respondent is asked about the date that the respondent is graduated and ready to work. In addition the respondent is asked how many months the respondent was unemployed since that date. The months of unemployment are subtracted from the amount of months to determine the total amount of work experience. The total amount of relevant work experience is asked with the following question: "How many months have you worked at your educational level?".

Categorical Variables

The operationalization's of the mandatory variables in the equations to test the hypotheses are explained in the previous sections. In addition a couple of categorical variables were included as control variables. The categorical variables are displayed in Table 1 and uses earlier in previous studies. Gender and race are primarily taken into account by previous studies

(Quintini, 2011), but also educational level, self-employment and Latin honours. These categories are added with a multiple-choice question in the questionnaire. On top of that the categorical variable region is added, mentioned as cause of over-education by Chevalier (2003).

Table 3

Categorical variables used in the OLS Regression and Logit Regression

Variable	Coding/Label	Variable	Coding/Label
Gender	0 = Woman 1 = Men	Race	0 = White 1 = Other
Latin honors	0 = No 1 = Yes	Self-employment	0 = No 1 = Yes
Education Level	0 = Bachelor 1 = Master 2 = Ph.D. / GZ -Psycholoog	Region	0= Non-restricted 1= Restricted

4. Results

The results section will present the empirical outcomes from the questionnaire to test the different hypotheses between the three labour market theories (HCT, JCM & AM) for OE-DPES graduates. The hypotheses are on the relationships between wage and over-education (H1a, H2a & H3a) and the difference between individual capabilities for OE individuals with CE individuals (H1b, H2b & H3b). As mentioned in the methodology section the hypotheses on wage are tested with the ORU-specification, estimated with a linear regression. The hypotheses on the difference between individual capabilities are tested with heterogeneity equation, estimated with a logistic regression. Before the empirical outcomes of the regressions are displayed the data is first examined and the amount of individual capabilities is reduced with a factor analysis.

4.1 Statistical Analysis

To properly examine the data, the dataset is cleaned with respect to missing values. Al together, 134 respondents filled in the questionnaire, of which 118 (88.1%) entirely and 16 (11.9%) respondents partly. Examining the 16 respondents displays that 12 of them did not continue after the first questions or reading the introduction. The remaining four respondents did not continue after the questions about the individual capabilities, which were the last questions. An improved survey design would be to split up the questions about individual capabilities to make it look less demanding. The remaining four respondents do have a value for the variable wage and match between education and job and still could be used for the analyses that do not demand the variables of individual capabilities. The respondents that were unemployed (9) at the moment the questionnaire was taken were guided to skip the questions about wage and the match between education and job and therefore could not be used. The logistic regression that demands the individual capabilities has a sample size of N = 110 and the linear regression that does not demand the individual capabilities has a sample size of N = 113. The sample sizes for both regressions are enough to test our hypotheses, but to draw substantive conclusions from our analysis some precautions are taken. The precautions taken are explained with the relevant analysis.

In all, there are maximally 113 usable respondents that answered the question: ”How do you evaluate your education with regard to your job?”. An amount of 63 respondents filled in that their job was appropriate with respect to their educational level and thus are quantified as CE. Seven of them filled in that their education was not enough for their job and thus are quantified as UE. The remaining 43 respondents filled in that their education was higher than needed for their job (N=37) or filled in that their education had no relation to their job (N=6) and thus are quantified as OE. From the 43 respondents that were quantified as OE 5 of them filled in that their education was a part of the job requirements. Due to contradictory answers these five OE respondents were transformed to invalid and thus no longer used for the statistical analyses. Chevalier recommends (2003) a difference between GOE and AOE caused by satisfaction with their job. There are 14 OE respondents that filled in that they were satisfied with their job and thus are transformed into an additional dummy of AOE. The remaining 24 respondents that were unsatisfied are transformed into an additional dummy GOE. An overview of the distributions can be found in table 4. Worth mentioning is the low amount of men (N=5) and the low amount of other ethnicity (N=2) in our sample size. There is evidence that the occurrence of over-qualification varies significantly by socio-demographic characteristics such as gender and ethnicity (Quintini, 2011). With the low representation of both men and other ethnicity it is not possible to draw substantive conclusions on the effect of gender and ethnicity on over-education. The low amount of men is a reasonable representation of the man / woman distribution within the study of Pedagogical and Educational Sciences (Universitairemasters, 2017).

Table 4

Frequency table variables

Variable	Frequency	Percentages
Respondent	134	
Finished	118	88.1%
Not Completed	16	11.9%
Match Job - Education	113	
Over-educated	43	38.1%
Genuine	24	
Apparent	14	
Invalid	5	
Correct educated	63	55.7%
Under educated	7	6.2 %
Gender		
Woman	126	96.2%
Man	5	3.8%
Ethnicity		
Dutch	129	98.5 %
Other Ethnicity	2	1.5 %
Education Level		
Bachelor	3	2.3%
Master	122	92.4%
Post-Master	7	5.3%
Education Direction		
DPES	112	89.6%
Different Bachelor	7	5.6%
Different Master	6	4.8%
Latin Honour		
None	105	84.0%
Yes	20	16.0%
Work Status		
Employed	114	91.2%
(Organization)	2	1.6%
Self-Employment	9	7.2%
Unemployed		
Region		
Non-Restricted	46	37.4%
Restricted	77	62.6%
Internship		
Yes	46	37.7%
No	76	62.3%

To get more insight in the data table 5 is added with descriptive statistics for the variables that are at least at the interval level. The first (Mean), second (Standard Deviation) and third (Skewness) moment of the distribution are displayed. The skewness of the variable informs about the symmetry of the distribution of those variables. If the skewness of a variable is less than -1 or greater than 1 the data is highly skewed (Brownmath, 2017). Experience, Months unemployed and Total Relevant Experience are positively skewed (right-skewed) and to the utmost extent highly skewed. The cause for the high positive skewness is that these variables can only be gained by time active on the labour market. Since the target group of the study is recent DPES the time on the labour market for these DPES graduates is quite short. Due to the low amount of months active on the labour market, Experience, Months Unemployed and Total Relevant Experience are unsymmetrical and skewed to the right. Now that the data is examined, the next step is to reduce the amount of individual capabilities to a useable amount for the analysis.

Table 5
Descriptive statistics (minimum interval level)

Variable	N	Minimum	Maximum	Mean	Std. Deviation	Skewness
Experience	112	0	136	34.52	26.62	0.939
Months Unemployed	123	0	32	2.54	4.95	3.319
Workhours	113	4	40	31.74	6.68	-1.273
Wage	113	1	12	7.26	1.97	-0.287
Total Relevant Experience	121	0	120	25.36	27.49	1.618

4.2 Factor Analysis

The amount of skills is reduced by performing an Exploratory Factor Analysis (EFA). An exploratory factor analysis (EFA) can be used to find interrelationships among large numbers of items to identify underlying dimensions (factors). The goal of performing the EFA with the capabilities is to find common capabilities that can be represented in underlying dimensions and produce a scale that would have general applicability. Scale purification is accomplished by factor analysing the different scores on the 21 capabilities. The method of principal axis factoring is chosen since it looks at the total variance in the database, with the primary concern maximizing variance estimation. Principal axis factoring creates a minimum amount of factors that will account for a maximum amount of variance. The factor analysis is performed with all respondents (N = 118) that completed the questionnaire. There are varying opinions about the required sample size. The minimum requirement for data reduction (with factor analysis), an amount of 4-5 * N per variable is proposed (Hair et al., 2014). Williams et al. (2010) propose a sample size of: 100 as poor, 200 as fair, 300 as good, 500 as very good, and 1000 or more as excellent. The total amount of respondents (118) is higher than both proposed sample sizes, but the sample size is tight. To test the sample adequacy Kaiser-Meyer-Olkin (KMO) test is used, a measure for the sample adequacy for factor analysis by measuring the proportion of variance among variables that might be common variance. For an acceptable sample adequacy the value of the test should be greater than 0.5 (Hair et al, 2014). On the KMO test our dataset scores a sample adequacy value of .771 (>0.5) and is acceptable. The Bartlett's test of significance is significant (p=0.000), which implies that there is equal variance in our data base.

In the previous paragraph we concluded that our database is suitable for an exploratory factor analysis and can be extracted with principal axis factoring. The next step is to reduce the large number of skills into underlying dimensions. Determining the amount of dimensions which are represented by factors is an iterative process, but basing it on eigenvalues (>1) is generally a good guide line. As Table 6 showed, in our data there are six factors that have an Eigenvalues >1 and a seventh factor that is slightly lower (0,962 < 1). Based on Eigenvalues there are six factors in the data, whereas the fifth (1.223 > 1) and sixth (1.106 > 1) are slightly higher than the benchmark. Because the fifth, sixth and seventh factor are close to the benchmark (>1) the amount of factors will be determined by examining the Pattern Matrix with each amount of factors. An oblique rotation (Promax) is chosen since it produces factors that are correlated, which produces more accurate results for research involving human behaviours and when the data does not meet a priori assumptions (Williams et al., 2010).

Table 6
Factor scores capabilities

Factor	Eigenvalues	% of Variance	Cumulative %
1	5.271	25.101	25.101
2	1.802	8.582	33.682
3	1.750	8.332	42.014
4	1.491	7.101	49.115
5	1.223	5.823	54.938
6	1.106	5.266	60.205
7	0.962	4.582	64.787

The goal is to identify the maximum amount of dimensions that have multiple item loadings. The cut-off for loadings is greater than 0.45 proposed by Hair et al. (2014) for our sample

(Appendix, section 7.3). The maximum amount of dimensions is identified by performing the factor analysis with seven factors: six factors with multiple item loadings and one factor with a single loading higher than our cut-off value. Since we aimed at multiple loadings, the item of the single loading (Service Orientation) is deleted from the analysis and only six factors remained. The following Pattern Matrix is retrieved from the EFA with six factors in SPSS.

Table 7
Pattern Matrix capabilities

	Factor					
	1	2	3	4	5	6
Systems Analysis	**0,780**					
Active Listening	**0,768**					
Instructing	**0,483**				0,446	
Critical Thinking		**0,744**				
Active Learning		**0,740**				
Negotiation			**0,804**			
Persuasion			**0,754**			
Management of Personnel Resources			**0,527**			
Judgment and Decisions				**0,737**		
Time Management				**0,520**		
Learning Strategies					**0,571**	
Science					**0,477**	
Monitoring						**0.632**
Systems Evaluation						**0.492**

Note: The following capabilities did not load: *Social Perceptiveness, Complex Problem Solving, Coordination, Reading Comprehension, Speaking, Writing.*

In all, six factors have multiple loadings higher than the cut-off value (≥0.450) in the Pattern matrix. Factor 1 has three loadings higher than the cut-off value: Systems Analysis (0.780), Active Listening (0.768) and Instructing (0.483) and can be associated with the construct of Communication skills. Factor 2 has two loadings higher than the cut-off value: Critical Thinking (744) and Active learning (0.740) and can be associated with the construct of Analytical thinking. Factor 3 has three factor loadings higher than the cut-off value: Negotiation (0.804), Persuasion (0.754) and Management of Personnel Resources (0.527) and can be associated with the construct of Conversational skills. Factor 4 has two factor loadings higher than the cut-off value: Judgment and Decisions (0.737) and Time Management (0.520) and can be associated with the construct of Situation identification. Factor 5 has two loadings higher than the cut-off value: Learning Strategies (0.571) and Science (0.477) and can be associated with the construct of Knowledge application. Factor 6 has two loadings higher than the cut-off value: Monitoring (0.632) and Systems Evaluation (0.492) and both items can be associated with the construct of Supervision skills. The following items are not represented in one of the factors: Complex Problem Solving, Coordination, Reading Comprehension, Speaking, Writing and the deleted item Social Perceptiveness.

Testing the internal consistency of these six constructs is problematic since four factors are a two-item construct and two factors are a three-item construct. This was expectable prior to the factor analysis, since there was no a priori or theoretical underlying structure. To test the internal consistency Cronbach's alpha is normally used, but with a two-item construct Cronbach's alpha almost always underestimates true reliability. Measuring the internal consistency of a two-item construct is therefore less meaningful and thus is generally advised to make measures more reliable by increasing the number of items (Eisinga, Grotenhuis & Pelzer, 2013). Furthermore, Sijtsma (2009) shows that Cronbach's alpha has little usage to identify the internal consistency, since it is only based on a single test. The internal consistency of the constructs is therefore tested with the Inter-Inter Correlation Matrix. The Inter-Inter Correlation matrix allows us to compare the correlations inside the items of a construct with the correlations outside the constructs. We denominate the identified constructs as reliable when the correlations inside the items of a construct are significantly higher than the correlations outside the constructs. To test the difference between the internal correlations and the external correlations within the factors the independent samples t-test is run. The internal correlations of the constructs ($\mu = 0.405$) are higher than the external correlations of the construct ($\mu = 0.194$). The internal correlations are significantly higher ($p=0.000$) according to the independent samples t-test and thereby the internal consistency is not violated.

Table 8

Independent Samples Test with correlations

Levene's Test for Equality for Variances		t-test for Equality of Means	
F	Sig	t	Sig (2-tailed)
2.470	0.118	7.090	0.000

With the EFA the capabilities retrieved from O'NET are reduced from 21 items to 6 underlying factors. Now that the internal consistency is not violated, the last step of the EFA is to produce a score of the factors that is applicable for the logistic regression. With SPSS there are various options to create a factor score for each DPES graduate. The factors are allowed to be correlated (due to the Promax Rotation) and therefore the regression scores are used. The benefit of regression scores is that it takes not only the correlation between the factors and between the factors and observed variables into account, but also the correlation among observed variables as well as the correlation among oblique factors (p. 4, DiStefano, 2009). The correlation scores are created with SPPS, by saving the 'factor scores'. The regression score for each DPES graduate displays how the individual scores on each capability in comparison with the other DPES graduates.

The goal of the EFA analysis was to reduce the capabilities for DPES graduates retrieved from O'NET. With the EFA, the capabilities retrieved from O'NET are reduced from 21 items to 6 factors, represented with regression scores. The six factors explain 61.57% of the variance altogether. To recall, the factors are: communication skills, analytical thinking, conversation skills, situation identification, knowledge application, and supervision skills.

4.3 Linear Regression

The first hypotheses about the consequences on wage for OE, CE and UE can be tested, now that the data is examined and the amount of capabilities is reduced. To test the hypotheses on wage the ORU-specification (see sector 3.2.1) is used with a linear regression using the Ordinary Least Squares (OLS) method. The OLS regression shows the relationship between the dependent variable and the independent variables. The coefficient obtained from the OLS regression displays how the dependent variable acts when there is an increase in the independent variable. In our model the dependent variable is wage and the OLS regression displays how the dependent variable acts on wage. With the OLS regression we could test the consequences of the match between education and job on wage and thereby test the corresponding hypotheses (H1a, H2a & H3a). In addition, control variables are added to the ORU-specification to control the relationship between the consequences of the match between education and job on wage. The maximum sample size of N = 113 is used, with all the respondents that answered the question between education and job (without the unemployed respondents). Currently unemployed respondents cannot be used since they did not have a score for wage and are left out of the analysis. The OLS regression can be used as long as the data does not reject the assumptions of the OLS regression. The assumptions of the OLS regression can be found in Appendix (section 7.4). None of the assumptions are violated when tested and there were no thorough reasons to delete respondents. It has to be noted that the assumptions are not perfectly met, but it seems that the reason is mostly caused by low sample size. No respondents are deleted since there were no thorough reasons and to keep as much respondents in the data as possible. Since the assumptions of the OLS are not violated, the hypotheses on wage are tested.

With the OLS-regression we can test the hypotheses about the difference in wage for OE individuals in comparison with CE individuals. According to the HCT "*the wage for OE individuals is equal in comparison with individuals that are CE*" and according to the JCM & AM "*the wage for OE individuals is lower in comparison with individuals that are CE".* The match between education and job level is made into dummies and one of the dummies has to be taken out and serve as a reference group. Since the focus of this research is on over-education the dummy for OE-DPES graduates serves as reference group and is taken out.

The hypotheses are tested with two different OLS regressions. The first specification is a basic Mincer Model (ORU-specification with dummies) derived from the HCT. The second specification is expanded with control variables.

The first regression formula is the ORU-specification with dummies, which results in the following regression formula and table:

$$\text{WAGE} = \beta_0 + \beta_1 D^{ReqEduc} + \beta_2 D^{UnderEduc} + \beta_3 Experience_i + \beta_4 Experience_i^2 + e$$

The second regression formula is with all the control variables is as follows:

$$\begin{aligned}\text{WAGE} = \beta_0 + \beta_1 D^{ReqEduc} + \beta_2 D^{UnderEduc} + \beta_3 Experience_i + \beta_4 Experience_i^2 \\ + \beta_5 Gender + \beta_6\, Region + \beta_7\, Latin\ Honors + \beta_8 Selfemployed \\ + \beta_9 Education\ Level + \beta_{10}\, Study\ direction + \beta_{11} WEP \\ + \beta_{12} Relevant\ Workexperience + \beta_{13}\, Workhours + e\end{aligned}$$

The first regression (table 9) displays that CE-DPES graduates earn more (β = 0.811, p=0.017) than OE-DPES graduates and the relationship is significant α = 0.05. In addition UE-DPES graduates tend to earn more than OE individuals (β = 0.201, p=0.769), however the relationship is not significant. The results tend to favour the hypotheses of JCM and AM and are contradictive to HCT. Experience is significant (β = 0.054, p=0.001) at α = 0.01 and also has a positive effect on wage. The ORU-specification is created to test the difference between over-education, correct education and under-education and is based on the HCT. The general success of the HCT is the strong predictive power of human behaviour instead of the accuracy of the HCT (Tan, 2014). The HCT will not work for every case, but it usually is a good predictor since its predictive power for wage is sufficiently large. In our sample size the

R^2 with the ORU-specification included has an explanatory power of 0.314 (Adj R^2 = 0.289) and the ORU-specification does not explain much of the variance.

Table 9
Regression ORU-specification (model only)

Variables	Beta	t	Sig.
(Constant)	5.182	13.473	0.000
Under education	0.201	0.294	0.769
Correct Education	0.811	2.432	0.017
Experience	0.054	3.348	0.001
Experience^2	0.000	-1.180	0.240

$*R^2 = 0.314$ (Adj. $R^2 = 0.289$)

In the second regression (table 10) the hypotheses are tested with the ORU-specification and control variables. The results of both regressions reveal that DPES graduates that work at the correct education level earn significantly more than DPES graduates that are OE for their job. The CE-DPES graduates have a positive beta (β=0.752, p=0.016) and is significant at the α = 0.01 level. The mean for OE DPES graduates on wage is 6.62 (~ €1655) and of CE DPES graduates 7.71 (~€1928). The wage for OE individuals is on average (€1655-€1928) / €1928 ≈ 14.16% lower than CE-DPES graduates. The results show that OE-DPES graduates earn significantly less than DPES graduates at the correct educational level. The hypothesis on wage according to the HCT: "*the wage for OE individuals is equal in comparison with individuals that are at the correct educational level"* is rejected by the results. The results are in line with the hypotheses of the JCM & AM *"the wage for OE individuals is lower in comparison with individuals that are at the correct educational level"* and is accepted by the results. The result that CE individuals earn more than OE individuals is in line with Sloane et al. (1999), McGuinness (2006) and other research to OE individuals.

Table 10
Regression ORU-specification (complete)

Variables	Beta	t	Sig.
(Constant)	-0.690	-0.464	0.644
Under education	0.030	0.055	0.957
Correct education	0.690	2.452	0.016
Experience	0.019	1.412	0.161
Experience^2	0.000	0.015	0.988
Gender	-0.299	-0.381	0.704
Region	-0.207	-0.760	0.449
Latin Honours	0.204	0.567	0.572
Self-employment	-1.157	-1.188	0.238
Education level	0.917	1.538	0.127
Study direction	-0.257	-0.620	0.537
WEP	-0.060	-0.221	0.825

Relevant Work experience	-0.002	-0.396	0.693
Workhours	0.172	8.214	0.000

R^2 = 0.639 (Adj. R^2 = 0.592)

In the first regression experience is significant (β=0.054, p=0.001) at α = 0.01 level, implying that more experience leads to a higher wage. In the second regression analysis the significance of experience has disappeared (β=0.161, p=0.161). There might be a spurious relationship between experience and wage since the significance disappeared. In the second regression the amount of workhours is added and is strongly significant (β=0.172, p=0.000) at α = 0.001 level and might cause the disappearance of significance for experience. No significant effect of experience on wage is contradicting to the HCT that predicts that experience is integrated with human capital and leads to a higher wage. The remaining variables are not significant in the second regression analysis. With the linear regression the hypotheses on wage (H1a, H2a & H3a) are tested. The results show that OE-DPES graduates tend to earn less than the other DPES graduates, confirming the hypotheses of JCM & AM and reject the hypothesis of HCT.

The last three hypotheses are about individual differences (H1b, H2b & H3b). So far the studies on over-education primarily focused on the effect of over-education on wage, whereas most often there is controlled for the difference in gender or ethnicity. Different studies, such as Hartog (2000), applied to investigate individual differences as an important cause for over-education. The difference within useful capabilities between occupations is a threshold to create a proxy for useful individual capabilities. By selecting a singular field of study with homogeneous occupations a singular proxy of useful individual capabilities for DPES graduates is identified. To test the last three hypotheses, the individual capabilities of OE-DPES graduates are compared with the individual capabilities of CE-DPES graduates and the hypotheses will be answered by the help of a logistic regression.

4.4 Logistic Regression

The differences between the three theories (HCT, JCM & AM) were with respect to wage and individual differences. The use of the logistic regression instead of the OLS regression is caused by the differences in measurement level between the variable wage and the variable over-education. Wage is measured at the interval level, while the variable over-education is categorical and dichotomous. In a linear regression the dependent variable (wage) is considered continuous whereas in the logistic regression the dependent variable is considered to be categorical. Since over-education is categorical a logistic regression is more suitable to test the hypotheses about individual differences. Since the assumptions are less strict for the logistic method (Appendix, section 7.5) and the assumptions are already tested for the OLS method no additional tests are performed.

The following hypotheses are tested: According to both HCT (H1b) and JCM (H2b) "*OE is not caused by individuals' capabilities (homogeneity)"* and according to the AM (H3b) "*OE is caused by individuals' capabilities (heterogeneity)".* To test the hypotheses about individual differences (H1b, H2b & H3b), the capabilities of each group of DPES graduates (OE, CE and UE) are compared to the other groups. The group of UE-DPES graduates only consists of seven respondents, which makes the comparison with UE-DPES graduates not feasible. Therefore just the comparison between the capabilities of the OE-DPES graduates (N=41) and CE-DPES graduates (N=63) is made. In addition, OE-DPES graduates are separated by satisfaction in two groups, GOE-DPES graduates (N=24) and AOE-DPES graduates (N=14), recommended in the study of Chevalier (2003). GOE and AOE-DPES graduates are compared separately to CE-DPES graduates. The goal of using the logistic regression is to test whether individual capabilities and other variables increase or decrease the chance of being OE. Associations between over-education and the individual capabilities were estimated using the odds ratios (ORs) as produced by the logistic regression procedure in SPSS. The significance of the model is assessed by the Omnibus Test of Model Coefficients, to check that our model is an improvement over the baseline model. The goodness of fit of our model is assessed by the Hosmer and Lemeshow test, that checks whether the observed event rates match the expected event rates in the subgroups. The overall goodness of model fit of our model is quantified using Nagelkerke R^2 (Levecque et al., 2017). In the model the individual capabilities are represented by the six factors obtained from the EFA: Communication Skills, Analytical Thinking, Conversation Skills, Situation Identification, Knowledge Application & Supervision Skills. In addition, the following control variables are added to the regression: Education level, Latin Honour, Region (restricted), work experience place (WEP), Study Direction & Relevant Experience.

The regression formula is as follows:

$Y_{\text{(Genuine \& Apparent) OE Individuals}} = p_i + e$

$\text{Log}\ [p_i/(1-p_i)] = \beta_0 + \beta_1\text{Education Level} + \beta_2\text{Latin Honor} + \beta_3\text{Regio} + \beta_4\text{WEP} + \beta_5\text{Study Direction} + \beta_6\text{Relevant Experience} + \beta_7\ \text{Communication Skills} + \beta_8\text{Analytical thinking} +$

β_9Conversation Skills + β_{10}Situation Identification + β_{11}Knowledge Application + β_{11}Supervision Skills

The groups of AOE (N= 14), GOE (N=24) and OE (N=41) are compared to CE (N=63). The results are displayed in the following table.

Table 11
Logistic regression of the capabilities

	Regression 1 *OE – CE*		Regression 2 *AOE - CE*		Regression 3 *GOE-CE*	
Omnibus Test of Model Coefficents	0.000		0.102		0.000	
Nagelkerke R Square	0.420		0.358		0.759	
Hosmer and Lemeshow Test	0.081		0.106		0.685	
	Odds Ratio	Sig.	Odds Ratio	Sig.	Odds Ratio	Sig
(Constant)	6.853	0.578	8.09E16	0.999	3.45E23	0.998
Communication Skills (F1)	1.803	0.272	1.654	0.515	4.352	0.184
Analytical Thinking (F2)	1.774	0.087	2.043	0.095	16.684	0.023
Conversation Skills (F3)	1.271	0.630	2.144	0.252	3.641	0.270
Situation Identification (F4)	0.635	0.328	0.372	0.118	0.124	0.079
Knowledge Application (F5)	0.472	0.168	0.282	0.072	0.093	0.091
Supervision Skills (F6)	0.871	0.741	0.481	0.18	0.462	0.385
Education Level	0.660	0.812	0.000	0.999	0.000	0.998
Latin Honor	3.037	0.170	2.715	0.379	0.8403	0.222
Region	0.710	0.517	0.541	0.492	3.387	0.217
WEP	0.970	0.956	0.919	0.915	0.092	0.091
Study Direction	0.085	0.040	0.309	0.352	0.000	0.998
Relevant Experience	0.942	0.000	0.952	0.057	0.776	0.004

In logistic regression 1 OE DPES graduates are compared to the reference group of CE DPES graduates. The probability ratio of the estimated model (Omnibus Test of Model Coefficients) is significant (0.000) and a Nagelkerke R^2 of 45.3%. The Hosmer and Lemeshow Test is significant, implying that the observed and actual amount of over-education in our data is significantly different than the predicted amount by our model. The results of logistic regression 1 displays that having relevant experience decreases the chance of being OE (OR= 0.942, p= 0.000). The odds ratio of 0.942 implies that for every month of relevant experience there is an decreases with 5,8 %[2] in the odds of being OE. An odds ratio above one implies an increase in the odds of being OE. With the result of experience decreasing the chance to be OE a remark needs to be made, since relevant experience increases every month being at a job at the right educational level. Also DPES graduates that have participated in or finished a different study direction are less OE (OR= 0.085, p=0.040). None of the individual capabilities are significant in the first regression when the OE individuals are compared to CE individuals. The absence of significance with individual capabilities implies that there is no significant distinction between the capabilities of OE DPES graduates and the capabilities of DPES graduates at the correct educational level in regression 1.

The grouping of Chevalier (2003) leads to a difference between GOE-DPES graduates and AOE-DPES graduates. Both of these OE-DPES graduates are compared to the CE-DPES graduates. Logistic regression '2' compares the AOE-DPES graduates to the CE-DPES graduates. Logistic regression '3' compares the GOE-DPES graduates to the CE-DPES graduates. In logistic regression '2' the probability ratio of the estimated model is not significant (p=0.102), with a Nagelkerke R^2 of 0.358. Our model is a weak predictor for AOE and does not score significantly better than a base model. The results can be influenced by the low number (N=14) of AOE-DPES graduates. In the model there is no significant variable that increases or decreases the chance that an individual is AOE instead of CE. The results show that AOE individuals are not significantly better or worse in any of the individual capabilities. Chevalier (2003) wrote as cause for AOE that an individual is restricted to a region wherein the spouse works. Being restricted to a region did not lead to a significant result for AOE (OR= 0.541, p= 0.492). The aim of the design of the questionnaire is to determine the capabilities of the individual and the match between education and job, AOE by our results is not caused by individual capabilities and is caused by other motives. To find these motives why someone is AOE a more qualitative design would be needed, which is not the aim of this thesis.

[2] (0.942-1)*100%=5.8%

In logistic regression '3' the OE DPES graduates that are unsatisfied with their match (GOE) are compared to CE-DPES graduates. The probability ratio of the estimated model is significant (0.000) with a high R^2 of 75.9%. The findings show that having relevant experience is significant (OR=0.776, p=0.004) and leads to a decreased chance of being OE, identical to the result of regression '1'. The same remark on the results of relevant experience as regression '1' needs to be made, relevant experience is not gained when being in a job below the educational level. Relevant experience can be gained by a job at the educational level or at a work experience place. In general the requirements for a work experience place will be lower and could be a good alternative method for DPES graduates for gaining experience. The downside of gaining relevant experience on a work experience place is that these are often underpaid (Julen, 2016).

From the individual capabilities the construct Analytical Thinking increases the chance (OR=16.684, p=0.023) of being GOE. A relatively high score on Analytical Thinking increases the chance of being OE and the skill Analytical Thinking seems not to be rewarded in the labour market. An alternative explanation is OE DPES graduates did not have the possibility to test their analytical thinking in practice within their job (and on their educational level). OE DPES graduates might assume that they possess the Analytical thinking skills, without the ability to test if this is the case. Besides Analytical Thinking, the constructs Conversation Skills and Communication Skills have an odds ratio above one (ns). This implies that GOE-DPES graduates score relatively better on these skills than CE-DPES graduates. The remaining skills Situation Identification, Knowledge Application and Supervision Skills have an odds ratio below one (ns). These results indicate that GOE-DPES graduates in general are not worse at the beneficial capabilities compared to CE-DPES graduates. The results of the individual capabilities on over-education are weak, with a positive relationship between Analytical thinking and over-education.

The goal of this thesis was to test the different testable predictions in the HCT, JCM & AM with respect to wage and capabilities for over-education. The hypotheses from the HCT, JCM & AM and results of our analyses are displayed in table 12. The results with the ORU-specification show that OE DPES graduates earn less than CE DPES graduates. The difference in wage is in line with JCM and AM, but contradicting to HCT. To answer whether individual capabilities influence the chance of being OE, the capabilities of GOE, AOE & OE DPES graduates are compared to CE-DPES graduates. GOE-DPES graduates have a significantly higher score on 'Analytical Thinking'. Analytical thinking seems to be negatively or less rewarded in the labour market in an occupation that is primarily focused on emotional intelligence. Our findings failed to reject the hypotheses of HCT (H1b) and the JCM (H2b) that assume homogeneity between DPES graduates. The hypothesis of AM about differences in individual capabilities is not accepted (H3b). The finding of the lower wage for OE individuals is in line with previous research. However the hypothesis about individual differences cannot be accepted, implying that JCM is favoured (table 12).

Table 12
Results-overview of hypotheses for HCT, JCM & AM

	Wage	Individual differences
Human Capital Theory	OE = CE = UE H1a: X	No effect H1b: ✓)
Job Competition Model	OE < CE > UE H2a: ✓	No effect H2b: ✓
Assignment Model	OE < CE > UE H3a: ✓	Cause OE H3b: X

Our findings are contradicting the results of Hartog (2000), Sloane et al. (1999) & McGuinness (2006) who favour the AM. However as Sloane (2003) points out in a latter study "*that perhaps too much weight is being attributed to the share of the unexplained component that is accounted for by differences in skill within models that attempt to control for such effects implicitly. As an alternative to such approaches, researchers should perhaps attempt to incorporate, where possible, some explicit measures of skill into their analysis*." In this analysis there are explicit measures of capabilities added, made possible by choosing a single graduation direction (Dutch Pedagogical Educational Science) of which graduates find jobs in a relative homogenous set of occupations. However those explicit skills have not proven to explain the difference between OE and CE DPES graduates. Prior research investigated larger sets of groups which did not allow identifying a single set of capabilities that are beneficial in a work field. To my knowledge this research is the first that has taken into account capabilities in a work field with respect to over-education. The downside of only choosing DPES graduates is the low generalizability in comparison to different graduation fields, more research to different graduation fields would be beneficial.

4.5 Discussion & Further Research

The results of the logistic regressions shows that individual capabilities do not explain why someone is OE, but these results are influenced by the low power (See Table 14). The power is determined by the sample size, effect size and the alpha level and are the interrelated components that determine the statistical power. The observed effect sizes are small to medium (displayed by Cohen's D), the sample sizes are on the low side, which in all results in a low power. Low power leads to an increased chance of a type II error, wherein the null hypothesis is not rejected where in fact the alternative hypothesis is true. Increasing the α to 0.10 which used as critical level in other behavioural studies will result in a higher power, but also increases the chance of a type I error (Bartlett et al., 2001). For further research an increased sample size would be beneficial to increase power, if the observed effect sizes (small) are similar to this research.

Table 14
Statistical Power Logistic Regression

	Regression 1: OE - CE				*Regression 2: GOE - CE*			
Variable	Cohen's D	Alpha	Sample	Power	Cohen's D	Alpha	Sample	Power
Communication Skills	-0.081	0.05	100	0.059	-0.320	0.05	82	0.298
Analytical Thinking	-0.232	0.05	100	0.205	-0.422	0.05	82	0.470
Conversation Skills	0.068	0.05	100	0.052	-0.060	0.05	82	0.045
Situation Identification	0.084	0.05	100	0.110	-0.035	0.05	82	0.036
Knowledge Application	0.182	0.05	100	0.230	-0.017	0.05	82	0.030
Supervision Skills	0.086	0.05	100	0.112	-0.339	0.05	82	0.323

The debate on causes for over-education is scattered by two obstacles. The first obstacle is the measurement to determine whether someone is OE. As mentioned in section 3.3, there are three measurements to determine whether someone is OE: the normative approach, self-declared approach and the statistic approach. The problem with the measurement is that different individuals are quantified as OE in different approaches (Battu et al. 2000). Hence when the OE individuals vary within each approach it is difficult to say something unambiguous. While the chosen method depends sometimes on the data available, for further research a unified method is helpful.

The second obstacle is the difficulty with quantifying individual's capabilities. Identifying the capabilities or a proxy for the capabilities has led to various discussions. Green et al. (2002) attempted to use the grades as proxy for individuals' capabilities, but they found little evidence in the UK labour market. A second method is the self-assessment method wherein respondents are asked directly to compare their total level of capabilities with the capabilities needed for respondent's job. Example from questions of the self-assessment method are: "I use many of my abilities in my current job", "My current job offers me sufficient scope to use my knowledge and skills" or "I would perform better in my current job if I possessed additional knowledge and skills" (Quintini, 2011). The downside of this method is displayed by a study of Allen and Van der Velden (2011) in data for the Netherlands, even graduates whose capabilities match their jobs report rather severe underutilization of their capabilities. Both methods prove to be ineffective and thereby this thesis has tried a new method. The method used in this paper is to identify the capabilities with O'NET and quantify each capability with a singular statement. The score on the statements determines the relative level of the respondent in the corresponding capability. The reliability of a singular statement for a capability is of course lower than using multiple statements, however singular statements reduced the amount of effort for the respondent. While this method is not applicable for large labour market research, it presents an additional method for quantifying individual's capabilities which is barely investigated in the literature. The downside of self-assessment of a capability is the subjective measurement. For further research the internal validity could be increased by attaining objective measurements of the capabilities.

The shortcoming but also the strength of this research is the chosen method to quantify individual's capabilities. While this is the first research that used actual capabilities, the downside is that only a single group of graduates is taken into account. DPES graduates are chosen because of a disturbed labour market wherein there are too many DPES graduates for the amount of jobs available (Julen, 2016). It is known that over-education is more common among Behavioural science, of which Pedagogical and Educational Science is part (Barone & Ortiz, 2010). Although this study clearly emphasizes the size of the problem of over-education among DPES graduates, the question arises whether this study can be extrapolated to the Dutch labour market. The last remark has to be made with respect to the different terminology in the literature for both the words 'over-education' and 'capabilities'. Different equivalent words are used across studies but main lines of the words are similar. To prevent ambiguity in this thesis only the words over-education and capabilities are used, but can therefore give a slightly different interpretation as originally intended by the authors.

For further research, as Sloane (2003) mentioned, it is important to include some explicit measures of the capabilities of the individual in the analysis. Only by opposing the capabilities of the individuals it can be encountered as (one of) the causes of over-education and thereby test the different predicted hypotheses of the HCT, JCM & AM. This thesis is a start of identifying and quantification of the capabilities using O'NET. The results are a falsification for JCM, but this study is also influenced by a low sample size. Commonality between the definition of over-education and a unified method for quantifying over-education will be useful for further research. The study of Quintini (2011) that combined the current literature in a general framework can be used as a good starting point.
The Job Competition Model is established in our research as best interpretation of the labour market for DPES graduates. According to the JCM, the wage is embedded in the job characteristics and not the characteristics of the individual (see section 2.3). The central element of JCM is that in the job the majority of workplace skills are acquired through on-the-job training as opposed to formal education (McGuinness, 2006). In a labour market with an abundance of DPES graduates the competition is fierce. DPES graduates compete for job opportunities instead of competing based on wage. DPES graduates are willing to accept a job wherein wage is less important than having a job. The interpretation by the JCM for the existence of over-education is the relative position on the labour queue, in which there is an abundance of educated workers. In the HCT with an abundance of educated workers, an individual would be less likely to participate in education as the benefit in wage is omitted. In the JCM, the same individual would now be more likely to participate in education as a defensive necessity, necessary to protect their place in the queue (McGuinness, 2006). The JCM emphasizes the importance of a person's relative position in the labour queue. If the individual is unsatisfied with their current relative position in the queue an upwards shift is possible by gaining job experience with a work experience place or additional education. Nonetheless, with an abundance of DPES graduates the competition is fierce and strengthens the relative power of firms in the labour market.

4.6 Practical explanations and implications

The JCM emphasizes that productivity of an individual is not determined by the individual itself, but by the job characteristics. Implying that to increase productivity in a job (i.e. job performance) an organization should look to these characteristics. Overall research failed to reveal a clean pattern between over-education and job performance (McKee-Ryan & Harvey, 2011). An explanation for the inconsistency between the relationship between over-education and job performance is that the relationship might be spurious, and might be as JCM suggested wherein job characteristics are decisive for productivity. While there is no evidence that the performance of OE workers is lower, over-education has various concerns for the individual itself and consequences for the organization. The commitment of workers in organizations is typically conceptualized as being comprised of three parts: affective (one's emotional attachment to the organization), continuance (the costs associated with the organization), and normative (a sense of duty or obligation to remain with the organization) (Maynard et al., 2006). An OE worker who perceives them self as over-educated (and under-paid) may feel less emotionally attached to the organization.

4.6.1. Over-education and job satisfaction

In research a clear negative pattern is found with the satisfaction on a job by over-educated workers (Jones et al., 2008). Over-educated workers have lower job satisfaction than those who were in graduate level jobs (Battu et al., 1999). Most studies focus on the comparison between OE workers and CE workers, while a little amount of studies make the comparison between OE workers and their co-workers. But over-education is associated with poorer job attitudes and negatively related to job satisfaction, work commitment, job involvement and internal work of motivation (Feldman & Turnley, 1995). OE workers have lower satisfaction due to the feeling of being relatively deprived in comparison with their counterparts that work at the correct educational level (according to the relative deprivation theory) (McKee-Ryan et al., 2009). The findings of Nabi (2003) display that over-educated individuals have lower opportunities to use skills and develop new ones in the context of their current jobs. Hence when over-educated compare themselves with their counterparts that work at the correct educational level they feel a slower development and the feeling of deprivation is enhanced. While both the AM and HCT emphasizes the qualities of the individual and the need for education, more focus should be on the side that JCM emphasizes namely the organizational side (demand side). When organizations create more opportunities to develop skills for over-educated workers and the feeling of deprivation is decreased, the disturbance for job satisfaction of OE workers is restored. As Nabi (p. 380, 2003) states in the implications of his findings: *"Although there was some relative room for improvement in terms of the supply side of skills, greater attention also needs to be placed on the demand side. This demand may be assisted through the creation of graduate level jobs in SMEs and many growing smaller companies are enhancing their graduate intake because they believe that graduates are better equipped than non-graduates. So creating demand for graduate skills is an important aspect of creating opportunities for skill use for over-educated graduates".*

Lower job satisfaction of OE workers also leads to an increase in some direct and indirect costs for the organizations. One of them is absenteeism of workers, which imposes various costs for the organization (e.g. lost output of the absent worker). A number of authors have considered the relationship between job satisfaction and absence. While a number of authors found a negative relationship between satisfaction and absence, the relationship between job satisfaction and absenteeism is also complex. It is difficult to distinguish between voluntary and involuntary absenteeism, hence the link between job satisfaction and absenteeism needs additional research (Jones et al. 2011). Related research that investigated mental health found that OE workers reported significantly more depression than CE workers (Dooley et al. 2000). Job satisfaction is also positively related to quitting. Dissatisfaction with the type of work was found to be the aspect most likely to lead to a worker leaving their job and being dissatisfied leads to a 35% higher probability of intending to quit than those reporting to be satisfied (Jones et al., 2008). Lower tenure of workers causes increased costs for the organisations, such as finding replacement or training new workers. OE workers are significantly more likely to be looking for new jobs or plan on looking for new jobs (Feldman & Turnley, 1995). In general the tenure for workers that are less satisfied is lower, however surprisingly no research links over-education to actual turnover behaviour (Maynard et al., 2006).

4.6.2. Over-education and hiring

The JCM explains the hiring process as a labour queue, wherein the individual with the most training cost will obtain the position. Being over-educated, as discussed in the previous paragraph, leads to lower opportunities for OE workers to increase their training costs. In our results no relationship could be found between the capabilities of the individual and over-education, implying that over-education is caused by any omitted variables in the regressions or by chance. While over-education is not caused by a lack of capabilities, the lack in opportunities to develop training costs leads to scarring. Thus, to avoid disadvantage themselves over-educated graduates should seek the opportunity to use and develop a wide range of skills to maintain their place in the labour que, Organizations should offer the opportunity for OE workers to negotiate for job enlargement and enrichment with their line manager or to move into graduate level jobs (Nabi, 2003). Both the OE worker and the employer are likely to benefit from finding ways in in which the fit between the person and the organisation can be maximized. Both recruitment and selection are important facets to determine the fit between the applicant and the organization. Research demonstrated that interviewers with limited training tend to overstate personal connection instead of capabilities. Application process assessments can be used to test different capabilities of the individual (Maynard et al. 2006). Bloemers (2017) wrote that often inefficient or even no assessments are used in the application process (p. 34) and that the application process is wrongly directed to personal connections instead of determining the capabilities (p. 45). These findings are in line with the results of the logistic regression that displays that there is no significant difference between the capabilities of OE DPES graduates and CE DPES graduates. Recruitment by organisations is crucial in hiring over-educated individuals to determine the fit between organisation and the person. The absence of a fit leads to lower job satisfaction and accompanying consequences. Possibilities for organisations is to find a different position (or even create a new one) for the OE worker within the company, to take advantage of his or hers capabilities (Maynard et al., 2006).

5. Conclusion

On balance, the evidence from our questionnaire among Dutch Pedagogical Educational Science graduates tends to support the *Job Competition* interpretation of the labour market, implying that the job characteristics are important for determining wage. Previous research (McGuinness, 2006; Battu et al., 1999; Hartog, 2000) tends to favour the *Assignment* interpretation of the labour market, wherein both individual and job characteristics are important. This research incorporated explicit measures of beneficial capabilities for DPES graduates, however none of the capabilities decreased the chance of being OE.

The three labour market theories (HCT, JCM and AM) offer certain insights and partial explanations for over-education, which are turned into a number of testable hypotheses. This thesis focused on how over-education among Dutch Pedagogical Education Science sits within the context of these three labour market theories by testing these testable hypotheses. The hypotheses are with the difference in wage between OE individuals and CE individuals and individual differences. This research focused on DPES graduates due to the bad labour market conditions wherein there are too many graduates in comparison with the amounts of jobs available (Julen, 2016). Over-education among the DPES graduates is widely represented (38.1% in our data) excluding the unemployed DPES graduates (6.7% in our data). The hypotheses on wage are tested with the ORU-specification with the dummy model in a linear regression. The results on wage show that OE DPES graduates earn significantly less than CE DPES graduates. These results confirm the hypotheses from the JCM and the AM, but are contradicting the HCT. The second hypotheses are about individual differences and if over-education is caused by different individual's capabilities. Useful individual capabilities for DPES graduates are derived from O'NET and thereby this is one of the first studies that included explicit measures for individual capabilities. The total amount of useful capabilities are reduced to six factors with a factor analysis to test these capabilities in a logistic regression. The results of the logistic regression show that over-education is not caused by a lack of certain capabilities and thereby the results are contradicting the AM but are in line with HCT and JCM. OE individuals even score significantly higher on one of the capabilities, namely Analytical Thinking. Our results show that OE DPES graduates earn significantly less but are not worse in any of the useful capabilities, favouring the JCM interpretation of the labour market.

6. 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.

Barone, C., & Ortiz, L. (2011). Overeducation among European University Graduates: a comparative analysis of its incidence and the importance of higher education differentiation. *Higher Education, 61 (335)*, 325-337.

Bartlett, J., Kotrlik, J., & Higgins, C. (2001). Organizational Research: Determining Appropriate Sample Size in Survey Research. *Information Technology, Learning, and Performance Journal, 19 (1)*, 43-50.

Battu, H., Belfield, C., & Sloane, P. (1999). Overeducation Among Graduates: a chort view. *Education Economics, 7 (1)*, 21-38.

Battu, H., Belfield, C., & Sloane, P. (2000). How Well Can We Measure Graduate Over-Education and its Effects? *National Institute Economic Review, 171 (1)*, 82-93.

Bloemers, W. (2017). *De nieuwe assesmentgids (42ste druk).* Antwerpen: Veen Bosch & Keuning uitgevers.

Brownmath. (2017, 9 24). *Shape Skewness*. Retrieved from www.brownmath.com: htpps:..brownmath.com/stat/shape.htm#Skew_CI

Brynin, M., & Longhi, S. (2009). Overqualification: Major or minor mismatch? *Economics of Education Review, 28 (1)*, 114-121.

Büchel, F., & Pollmann-Schult, M. (2004). Overeducation and human capital endownments. *International Journal of Manpower, 25 (2)*, 150-167.

Büchel, F., & van Ham, M. (2003). Overeducation, regional labour markets, and spatial fexibility. *Journal of Urban Economics, 53 (3)*, 482-493.

CBS. (2017). *Beroepsbevolking; behaalde onderwijs naar persoonskenmekren 2001-2012*. Retrieved from http://statline.cbs.nl/StatWeb/publication/?VW=T&DM=SLNL&PA=71822NED&D1=0&D2=a&D3=a&D4=0-1,4&D5=a&D6=0&D7=2,l&HD=130926-1540&HDR=T,G3,G5,G6,G1&STB=G2,G4

Chevalier, A. (2003). Measuring Over-Education. *Economica, 70 (279)*, 509-531.

Chiswick, B., & Miller, P. (2009). The international transferability of immigrants' human capital. *Economics of Education Review, 28 (2)*, 162-169.

DiStafano, C., Zhu, M., & Mindrila, D. (2009). Understanding and Using Factor Scores: Considerations for the Applied Researcher. *Practical Assesment Research Evaluation, 14 (20)*, 1-11.

Dolton, J., & Silles, M. (2008, 27 (2)). The effects of over-education on earnings in the graduate labour market. *Economics of Education Review*, 125-139.

Dolton, P., & Vignoles, A. (2000). The incidence and effects of overeducation in the U.K. graduate Labour market. *Economics of Education Review, 19 (2)*, 179-798.

Dooley, D., Prause, J., & Ham-Rowbottom, K. (2000)). Underemployment and Depression: Longitudinal Relationships. *Journal of Health and Social Behavior, 41 (4)*, 421-436.

Eisinga, R., Grotenhuis, M., & Pelzer, B. (2013). The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? *International Journal of Public Wealth, 58 (4)*, 637-642.

Feldman, D. (1996). The Nature, Antecedents and Consequences of Underemployment. *Journal of Management, 22 (3)*, 385-407.

Feldman, D., & Turnley, W. (1995). Underemployment Among ecent Business College Graduates. *Journal of Organizational Behavior, 16 (6)*, 691-706.

Frei, C., & Sousa-Poza, A. (2012). Overqualification: permanent or transitory? *Applied Economics, 44 (14)*, 1837-1847.

Green, F., & McIntosh, S. (2007). Is there a genuine under-utilization of skills amongst the over-qualified? *Applied Economics, 39 (4)*, 427-439.

Green, F., McIntosch, S., & Vignoles, A. (2002). The Utilization of Education and Skills: Evidence from Britain. *The Manchester School, 70 (8)*, 792-811.

Groot, W. (1996). The incidence of, and returns to overeducation in the UK. *Applied Economics, 28 (10)*, 1345-1350.

Groot, W., & van den Brink, H. (2000). Overeducation in the labor market: a meta-analysis. *Economics of Education Review, 19 (2)*, 149-158.

Hair, J., Black, W., Tatham, R., Anderson, R., & Babin, B. (2014). *Multivariate Data Analysis (7th edition).* Pearson.

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

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

Hill, R., Griffiths, W., & Lim, G. (2011). *Principles of Econometrics (Fourth edition).* John Wiley & Sons Ltd.

Jones, M., Jones, R., Latreille, P., & Sloane, P. (2008). Training, Job Satisfaction, and Workplace Performance in Britain: Evidence from WERS 2004. *IZA Discussion Papers, 3677*.

Julen, J. (2016, 11 12 16 April, April 11, 14, 15). Werken voor een paar honder euro: noodzaak of uitbuiting. *Trouw*, Retrieved from: https://www.trouw.nl/home/werken-voor-een-paar-honderd-euro-noodzaak-of-uitbuiting-~abcc89e1/.

Lemieux, T. (2006). The "Mincer Equation" Thirty Years after Schooling, Experience, and Earnings. *Jacob Mincer A Pioneer of Modern Labor Economics*, 127-145.

Levecque, K., Anseel, F., De Beuckelaer, A., Van der Heyden, J., & Gisle, L. (2017). Work organization and mental health problems in PhD students. *Research Policy, 46 (4)*, 868-879.

Linsley, I. (2005). Causes of overeducation in the Australian market. *Journal of Labour Economics, 8 (2)*, 121-143.

Maynard, D., Joseph, T., & Maynard, A. (2006, 27 (4)). Underemployment, Job Attitudes, and Turnover Intentions. *Journal of Organizational Behavior*, 509-536.

McGuinness, S. (2006). Overeducation in the Labour Market. *Journal of Economic Surveys*, 387-418.

Mckee-Ryan, F., & Harvey, J. (2011). "I Have a Job, But ..." A review of underemployment. *Journal of management, 37 (4)*, 1-36.

Mincer, J. (1974). *Schooling, Experience, and Earnings.* New York: Columbia University Press.

Muysken, J., & ter Weel, B. (1999). Overeducation, job competition and unemployment. *Maastricht Economic Research Institute in Innovation and Technology*, 1-29.

Nabi, G. (2003). Graduate employment and underemployment: opportunity for skill use and career experiences amongst recent business graduates. *Education + Training, 45 (7)*, 371-382.

NVO. (2017, September 13). *Werkvelden*. Retrieved from Nederlandse Vereniging van Pedagogen en Onderwijskundigen: https://www.nvo.nl/

Olaniyan, D., & Okemakinde, T. (2008). Human Capital Theory: Implications for Education Development. *Pakistan Journal of Social Sciences, 5 (5)*, 479-783.

Peng, C., Lee, K., & Ingersoll, G. (2002). An Introduction to Logistic Regerssion Analysis and Reporting. *The Journal of Educational Research, 96 (1)*, 3-14.

Peterson, N., Mumford, M., Borman, W., Jeanneret, P., Fleishman, E., Levin, K., . . . Dye, D. (2001). Understanding work using the occupational information network (O*NET): Implications for practis and research. *Personnel Psychology, 54 (2)*, 451-492.

Polachek, S. (2008). Earnings Over the Life Cycle: The Mincer Earnings Function and Its Applications. *Foundations and Trends in Microeconomics, 4 (3)*, 165-275.

Quintini, G. (2011). Over-Qualfied or Under-Skilled: A Review of Existing Literature. *OECD Social, Employment and Migration Working Papers, 121*.

Scurry, T., & Blenkinsopp, J. (2011). Under-employment among recent graduates: a review of the literature. *Personnel Review, 40 (5)*, 643-659.

Sijtsma, K. (2009). Over misverstanden rond Cronbachs alfa en de wenselijkheid van alternatieven. *De Psycholoog, 44*, 561-567.

Sloane, P., Battu, H., & Seaman, P. (1999). Overeducation, undereducation and the British labour market. *Applied Economics, 31 (11)*, 1437-1453.

Tan, E. (2014). Human Capital Theory: A holistic Criticism. *Review of Educational Research, 84 (3)*, 411-445.

Tsai, Y. (2010). Returns to overeducation: A longitudinal analysis of the U.S. labor market. *Economics of Education Review, 29 (4)*, 606-617.

Universitairemasters. (2017, 10 11). *Aantal studenten gedrag en maatschappij*. Retrieved from Universitairemasters: http://universitairemasters.nl/aantal-studenten/gedrag-en-maatschappij

Verhaest, D., & Omey, E. (2006). The Impact of Overeducation and its Measurement. *Social Indicators Research, 77 (3)*, 419-448.

Williams, B., Onsman, A., & Brown, T. (2010). Exploratory factor analysis: a five-step guide for novices. *Journal of Emergency Primary Health Care, 8 (3)*, 1-13.

7. Appendix

7.1 Questionnaire

Start of Block: Introductie

Q1 Hartelijk dank dat u mij wilt helpen met het onderzoek!

Het invullen van deze enquête duurt maximaal 10 minuten.
Probeer de vragen zo eerlijk mogelijk te beantwoorden. De gegevens zullen vertrouwelijk en anoniem worden behandeld.

In deze enquête zijn we benieuwd naar de arbeidskansen en moeilijkheden daarin voor afgestudeerden in Pedagogische Wetenschappen.
Er wordt gevraagd naar het verloop van uw carrière en uw huidige baan, waardoor het eventueel handig kan zijn om uw CV bij de hand te houden.
Aan het eind van het onderzoek worden er vragen gesteld over uw persoonlijkheid.
Probeer altijd een antwoord in te vullen, ook als u twijfelt. Denk eraan dat er géén goede of foute antwoorden zijn.

Heeft u vragen over de enquête of over het onderzoek? Dan kunt u altijd contact met mij opnemen door mij een email te sturen naar: lj.postma@gmail.com

De enquête kan ook worden ingevuld op een telefoon. Als u de telefoon dwars houdt, is de enquête makkelijker in te vullen.

Hartelijk dank,
Lars Postma

End of Block: Introductie

Start of Block: Algemene vragen

Q2 Wat is uw geslacht?

○ Man (1)

○ Vrouw (2)

Q3 Wat is uw etniciteit?

○ Nederlands (1)

○ Anders namelijk... (2) ____

Q4 Wat is de hoogst genoten opleiding die u heeft afgerond?

- Bachelor (1)
- Master (2)
- Ph.D. (3)
- GZ psycholoog (4)
- Anders, namelijk (5) ____

Display This Question:

If Wat is de hoogst genoten opleiding die u heeft afgerond? != Bachelor

Q5 Kunt u aangeven in welke richting de bachelor en master behaald is?

- Bachelor Pedagogische Wetenschappen (1)
- Bachelor Psychologie (2)
- Anders, Bachelor behaald in (3) ____
- Master Pedagogische Wetenschappen (4)
- Master Psychologie (5)
- Anders, Master behaald in (6) ____

Display This Question:

If Wat is de hoogst genoten opleiding die u heeft afgerond? = Bachelor

Q6 Kunt u aangeven in welke richting de bachelor behaald is?

- Pedagogische Wetenschappen (1)
- Psychologie (2)
- Anders, namelijk (3) ____

Q7 Heeft u een judicia behaald tijdens uw opleiding? (cum laude of hoger)
Zoja, kunt u hieronder aangeven welke judicia u heeft behaald voor welke fase.

	Niet van toepassing (1)	Cum laude (2)	Summa cum laude (3)
Bachelor (1)	☐	☐	☐
Master (2)	☐	☐	☐

JS

Q8 Sinds welk moment bent u klaar met uw studie en op zoek naar deeltijds of voltijds betaald werk?

	Maand	Jaar
1 (1)	▼ Januari (1) ... December (12)	▼ 2000 (1) ... 2017 (18)

Q9 Hoeveel maanden bent u in totaal werkloos geweest sinds dat u afgestudeerd bent en op zoek naar deeltijds of voltijds betaald werk was?

Aantal maanden: (1)	

Q10 Heeft u op dit moment een baan?

- ○ Ja, ik werk voor een organisatie (1)
- ○ Ja, ik ben eigen baas (2)
- ○ Nee (3)

Q11 Bent u gebonden aan een specifieke regio of specifieke regio's voor een baan?

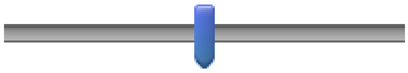
- ○ Ja (1)
- ○ Nee (2)

End of Block: Algemene vragen

Start of Block: Vragen over uw huidige baan

Display This Question:

If Heeft u op dit moment een baan? != Nee

Q12 Hoeveel uur werkt u gemiddeld betaald in de week? Aantal uren: (1)	

Display This Question:

If Heeft u op dit moment een baan? != Nee

Q13 Wat is uw netto-salaris per maand van uw huidige betaalde baan?

- ◯ Minder dan 499 (1)
- ◯ €500 - €749 (2)
- ◯ €750 - €999 (3)
- ◯ €1000 - €1249 (4)
- ◯ €1250 - €1499 (5)
- ◯ €1500 - €1749 (6)
- ◯ €1750 - €1999 (7)
- ◯ €2000 - €2249 (8)
- ◯ €2250 - €2499 (9)
- ◯ €2500 - €2749 (10)
- ◯ €2750 - €2999 (11)
- ◯ Meer dan €3000 (12)

Display This Question:

If Heeft u op dit moment een baan? != Nee

Q14 Is uw hoogste opleiding behaald een vereiste voor het beoefenen van uw huidige baan?

- ◯ Ja, de hoogste opleiding behaald was een vereiste (1)
- ◯ Nee (2)
- ◯ Weet niet (3)

Display This Question:

If Heeft u op dit moment een baan? != Nee

Q15 Hoe evalueert u uw hoogst behaalde opleiding met betrekking tot uw huidige baan?

- ◯ Mijn hoogste behaalde opleiding is niet voldoende. (1)
- ◯ Mijn hoogste behaalde opleiding past bij mijn baan. (2)
- ◯ Mijn hoogste behaalde opleiding is hoger dan nodig voor mijn baan. (3)
- ◯ Voldoende of niet voldoende is niet aan de orde, mijn hoogste opleiding heeft geen betrekking op mijn baan. (4)

Display This Question:

If Heeft u op dit moment een baan? != Nee

Q16 Hoe tevreden bent u met de match tussen uw huidige baan en uw opleiding?

	Zeer ontevreden (1)	Ontevreden (2)	Noch ontevreden, noch tevreden (3)	Tevreden (4)	Zeer Tevreden (5)
(1)	◯	◯	◯	◯	◯

End of Block: Vragen over uw huidige baan

Start of Block: Ervaring

Q17 Heeft u na uw studie ervaring opgedaan op een werkervaringsplek of stageplek?

Zoja, kunt u aangeven hoeveel maanden dit is geweest, onafgezien van het arbeidspercentage.

	Werkervaringsplek of stageplek		Aantal maanden:
	Ja (1)	Nee (2)	

(1)	○	○	▼ 0 (1) ... 24 (25)

Q18 Hoeveel maanden hebt u "op niveau" gewerkt, onafgezien van het arbeidspercentage?

Let op, een werkervaringsplek of stageplek telt NIET mee.

1 (1)	

Q22 In welke mate heeft de kennis verkregen op de opleiding bijgedragen aan het krijgen van een baan?

○ Niet (1)

○ In beperkte mate (2)

○ In aanzienlijke mate (3)

End of Block: Ervaring

Start of Block: Vragen over karakteristieken van de persoon

Q19 De volgende vragen gaan over uw persoonlijkheid, dit zijn tevens de laatste vragen.

	Zeer mee oneens (1)	Oneens (2)	Noch mee oneens, noch mee eens (3)	Eens (4)	Zeer mee eens (5)
Ik kan goed begrijpen wat andere mensen proberen te vertellen. (1)	○	○	○	○	○
Ik vind het moeilijk om iets uit te leggen. (2)	○	○	○	○	○
Ik ben actief op zoek om mensen te helpen. (3)	○	○	○	○	○
Ik ben goed in begrijpend lezen. (4)	○	○	○	○	○
Ik vind het moeilijk om sociale omstandigheden in te schatten. (5)	○	○	○	○	○
Ik vind het moeilijk om te zeggen wat ik bedoel. (6)	○	○	○	○	○
Ik ben goed in het monitoren en beoordelen van mijn eigen acties. (7)	○	○	○	○	○
Ik ben goed in het monitoren en beoordelen van andermans acties. (8)	○	○	○	○	○
Ik gebruik wetenschappelijke regels en methoden om problemen op te lossen. (9)	○	○	○	○	○

Q20

	Zeer mee oneens (1)	Oneens (2)	Noch mee oneens, noch mee eens (3)	Eens (4)	Zeer mee eens (5)
Ik heb moeite om de juiste keuze te maken. (1)	○	○	○	○	○
Ik gebruik een passende methodiek bij het leren van nieuwe dingen. (2)	○	○	○	○	○
Ik gebruik een passende methodiek bij het onderwijzen van nieuwe dingen. (3)	○	○	○	○	○
Ik kan op papier zetten wat ik bedoel. (4)	○	○	○	○	○
Ik kan goed nieuwe informatie opnemen en direct toepassen. (5)	○	○	○	○	○
Ik ben goed in het logisch redeneren. (6)	○	○	○	○	○
Ik vind het moeilijk om in te schatten hoe lang iets duurt. (7)	○	○	○	○	○
Ik kan mij aanpassen aan een nieuwe situatie. (8)	○	○	○	○	○

Q21

	Zeer mee oneens (1)	Oneens (2)	Noch mee oneens, noch mee eens (3)	Eens (4)	Zeer mee eens (5)
Ik vind het moeilijk om iemand te overtuigen. (1)	○	○	○	○	○
Ik kan complexe vraagstukken bestuderen en eventuele oplossingen implementeren. (2)	○	○	○	○	○
Ik zie welke acties nodig zijn om mezelf of andere prestaties te verbeteren. (3)	○	○	○	○	○
Ik kan mensen samenbrengen en tot overeenstemming laten komen. (4)	○	○	○	○	○
Op het werk kan ik collega's aansturen, ontwikkelen en motiveren. (5)	○	○	○	○	○
Ik ben bewust van de gevolgen en consequenties van mijn eigen en andermans gedrag. (6)	○	○	○	○	○

End of Block: Vragen over karakteristieken van de persoon

7.2 Skill list O’NET + Questions Questionnaire

Skill	Question:
1. Active Listening	Ik kan goed begrijpen wat andere mensen proberen te vertellen.
2. Instructing *	Ik vind het moeilijk om iets uit te leggen.
3. Service Orientation	Ik ben actief op zoek om mensen te helpen.
4. Reading Comprehension	Ik ben goed in begrijpend lezen.
5. Social Perceptiveness *	Ik vind het moeilijk om sociale omstandigheden in te schatten.
6. Speaking *	Ik vind het moeilijk om te zeggen wat ik bedoel.
7. Monitoring	Ik ben goed in het monitoren en beoordelen van mijn eigen acties. Ik ben goed in het monitoren en beoordelen van andermans acties.
8. Judgment and Decision Making *	Ik heb moeite om de juiste keuze te maken
9. Learning Strategies	Ik gebruik een passende methodiek bij het leren van nieuwe dingen. Ik gebruik een passende methodiek bij het onderwijzen van nieuwe dingen.
10. Writing	Ik kan op papier zetten wat ik bedoel.
11. Active Learning	Ik kan goed nieuwe informatie opnemen en direct toepassen.
12. Critical Thinking	Ik ben goed in het logisch redeneren.
13. Time Management *	Ik vind het moeilijk om in te schatten hoe lang iets duurt.
14. Coordination	Ik kan mij aanpassen aan een nieuwe situatie.
15. Persuasion *	Ik vind het moeilijk om iemand te overtuigen.
16. Complex Problem Solving	Ik kan complexe vraagstukken bestuderen en eventuele oplossingen implementeren.
17. Systems Evaluation	Ik zie welke acties nodig zijn om mezelf of anderen prestatie te verbeteren.
18. Management of Personnel Resources	Op het werk kan ik collega's aansturen, ontwikkelen en motiveren.
19. Systems Analysis	Ik ben bewust van de gevolgen en consequenties van mijn eigen en andermans gedrag.
20. Science	Ik gebruik wetenschappelijke regels en methoden om problemen op te lossen.
21. Negotiation	Ik kan mensen samenbrengen en tot overeenstemming laten komen.

- = Reversed questions

7.3 Critical value factor analysis (Hair et al. 2014)

TABLE 2 Guidelines for Identifying Significant Factor Loadings Based on Sample Size

Factor Loading	Sample Size Needed for Significance[a]
.30	350
.35	250
.40	200
.45	150
.50	120
.55	100
.60	85
.65	70
.70	60
.75	50

[a] Significance is based on a .05 significance level (α), a power level of 80 percent, and standard errors assumed to be twice those of conventional correlation coefficients.

Source: Computations made with SOLO *Power Analysis*, BMDP Statistical Software, Inc., 1993.

7.4 Assumptions linear regression

The assumptions are given in the following table (Hill & Griffiths, 2011, p. 173):

ASSUMPTIONS OF THE MULTIPLE REGRESSION MODEL

MR1. $y_i = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_K x_{iK} + e_i,\ i = 1, \ldots, N$

MR2. $E(y_i) = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_K x_{iK} \Leftrightarrow E(e_i) = 0$

MR3. $\mathrm{var}(y_i) = \mathrm{var}(e_i) = \sigma^2$

MR4. $\mathrm{cov}(y_i, y_j) = \mathrm{cov}(e_i, e_j) = 0 \quad (i \neq j)$

MR5. The values of each x_{ik} are not random and are not exact linear functions of the other explanatory variables

MR6. $y_i \sim N[(\beta_1 + \beta_2 x_{i2} + \cdots + \beta_K x_{iK}), \sigma^2] \Leftrightarrow e_i \sim N(0, \sigma^2)$

7.5 Assumptions logistic regression

The assumptions are given in the following table (Hill & Griffiths, 2011, p. 173)

ASSUMPTIONS OF THE SIMPLE LINEAR REGRESSION MODEL-I

- The mean value of y, for each value of x, is given by the linear regression function

$$E(y|x) = \beta_1 + \beta_2 x$$

- For each value of x, the values of y are distributed about their mean value, following probability distributions that all have the same variance,

$$\mathrm{var}(y|x) = \sigma^2$$

- The sample values of y are all uncorrelated and have zero covariance, implying that there is no linear association among them,

$$\mathrm{cov}(y_i, y_j) = 0$$

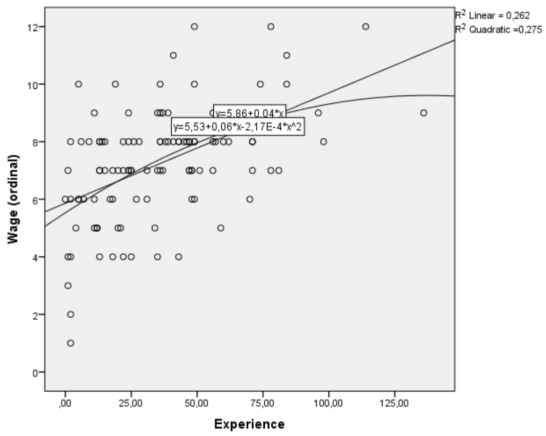
  This assumption can be made stronger by assuming that the values of y are all statistically independent.
- The variable x is not random and must take at least two different values.
- (*optional*) The values of y are normally distributed about their mean for each value of x,

$$y \sim N\left[(\beta_1 + \beta_2 x), \sigma^2\right]$$

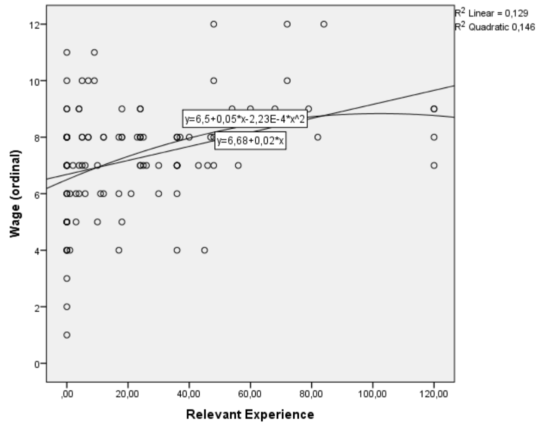
7.6 Testing assumptions linear regression

To test the linearity a linear trend-line and a quadratic trend-line are added to the plot-graphs with the dependent and the various independent variables. In the plot-graphs there is no reason to assume non-linearity, however there might be potential outliers. Based on the plot graphs there is no reason to reject the assumption on linearity. In general the quadratic line moves similar as the linear line, however in the graph between experience and wage there is probably an influential outlier. Causing the quadratic line to shift downwards.

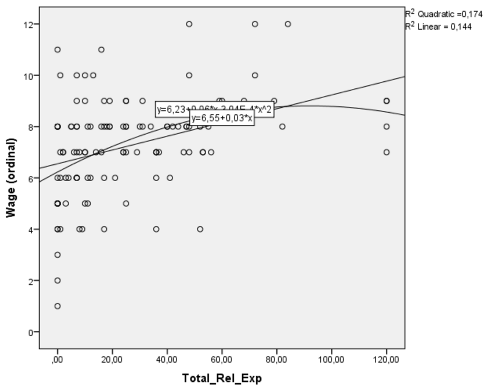
Experience – Wage



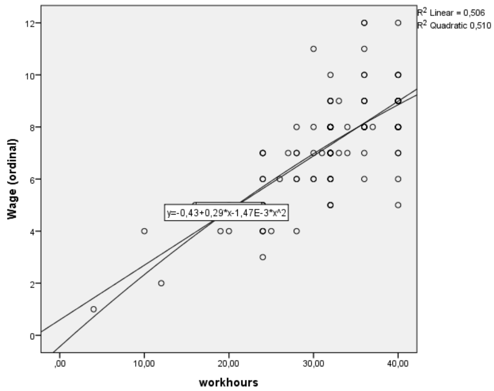
Relevant Experience - Wage



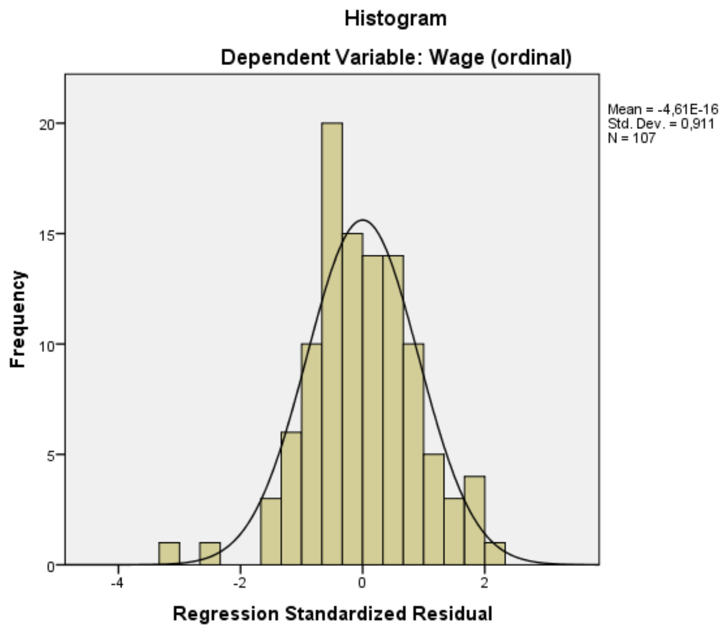
Total Relevant Experience – Wage



Workhours – Wage



Outliers will be later detected and studied to identify if they are influential cases. To test the independence of residuals the Durbin-Watson test is used, a value close to 2 is desirable and indicate that the residuals are dependent. The value of the Durbin-Watson is 2.156 and thereby we still assume independence of residuals. The residuals also need to be normal distributed, to identify normal distribution of the residuals the histogram of the Standardized residuals is used.



The residuals are right-skewed, but not far enough to reject the assumption about the normal distribution of the residuals. To test the assumption of homoscedasticity the standardized predicted residuals are plotted against the regression studentized residuals. In the plotted graph there seems to be equal variance in the residuals. To test if there is no Multi-Collinearity we look at the VIF-value.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B		Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	4,499	1,694		2,656	,009	1,133	7,866		
	UE	,022	,705	,003	,031	,976	-1,380	1,424	,724	1,382
	RE	,605	,351	,153	1,724	,088	-,092	1,302	,724	1,382
	Experience	,030	,007	,410	4,417	,000	,017	,044	,659	1,518
	Dummy Direction	,103	,555	,016	,186	,853	-,999	1,205	,776	1,288
	D_WEP	-,532	,346	-,131	-1,539	,127	-1,220	,155	,785	1,273
	Regio	,135	,330	,034	,410	,683	-,521	,792	,834	1,199
	Selfemployed	-2,790	1,208	-,192	-2,310	,023	-5,190	-,390	,823	1,215
	Judicia	-,142	,454	-,027	-,313	,755	-1,044	,760	,764	1,310
	Education_Level	,980	,754	,114	1,300	,197	-,519	2,480	,731	1,368
	Ethnicity	-1,868	1,716	-,091	-1,088	,279	-5,278	1,542	,808	1,238
	Gender	-,351	1,000	-,029	-,351	,726	-2,339	1,637	,807	1,238
	F1	-,167	,275	-,074	-,607	,545	-,713	,379	,380	2,632
	F2	-,219	,214	-,099	-1,026	,308	-,644	,205	,607	1,648
	F3	,142	,298	,066	,475	,636	-,451	,735	,298	3,356
	F4	,535	,263	,225	2,039	,044	,014	1,057	,463	2,158
	F5	,428	,275	,173	1,559	,122	-,118	,974	,459	2,177
	F6	,164	,265	,070	,618	,538	-,363	,691	,447	2,238
	F7	,087	,219	,034	,398	,692	-,349	,523	,797	1,254

a. Dependent Variable: Wage (ordinal)

A high VIF-value indicate that two or more variables correlate. In the ORU-specification there is experience and a generated quadratic form of experience, but the correlation between these variables is high (>10). Multi-collinearity has to be checked in order to protect when you estimate the independent effect of two variables which happen to be correlated by chance. Since quadratic form is a higher-order term of experience this is not by chance multi-collinearity is not a problem. We can conclude that the assumptions of the OLS-regression are not violated, but because we have a low sample size there might be influential cases. To identify these cases we look at Cook's D and the LEVER value. Cook's D indicate both the difference between beta's estimated with and without the individual case. The critical point for Cook's D is 4/n, whereas n stands for the amount of respondents (106). The critical value will be 0.0377 (4/106). The LEVER value measures the distance of the value of an individual case on a predictor to the center of the values of the other individual cases on that predictor. The critical point for the LEVER value is 2*p/n, whereas p stands for the amount of items in the analysis and n stands for amount of respondents. The OLS regression with the most amount of items has 19 items and therefore the most cautious LEVER value for our OLS regression is 0.3585 (2*19/106). In total there are twelve respondents that score above the critic value of either the LEVER value of Cook's D value. Closer examination of the respondents reveal that nine of the twelve are part of a minority category in a variable and thereby have an extreme X-value and high leverage. These twelve respondents are exceptions but there are no thorough reasons to delete one of those twelve respondents. Deletion of respondents is a bit restrained caused by the low sample size. All of this results that none of the assumptions of the OLS regression are violated and the OLS regression can be performed. It has to be noted that the assumptions are not perfectly met, but it seems that the reason is mostly caused by a low sample size.

7.7 Syntax SPSS

```
*Creating Dummies Over-, Req-, Under- Education.

RECODE Q15 (1=1) (ELSE=0) INTO UnderEducation.
VARIABLE LABELS  UnderEducation 'UE'.
EXECUTE.

RECODE Q15 (3 thru 4=1) (ELSE=0) INTO OverEducation.
VARIABLE LABELS  OverEducation 'OE'.
EXECUTE.

RECODE Q15 (2=1) (ELSE=0) INTO ReqEducation.
VARIABLE LABELS  ReqEducation 'RE'.
EXECUTE.

DO IF (Q14 = 1).
RECODE OverEducation (1=SYSMIS).
END IF.
EXECUTE.

*Create difference apparant education and genuine over-education.
DO IF (Satisfaction <= 3).
RECODE A_OE (1=0).
END IF.
EXECUTE.

DO IF (Satisfaction >= 4).
RECODE G_OE (1=0).
END IF.
EXECUTE.
```

*Extra recode for incorrect over-education

```
RECODE Q15 (3 thru 4=1) (ELSE=0) INTO OE2.
VARIABLE LABELS  OE2 'OE2'.
EXECUTE.

DO IF (Q14 = 1).
RECODE OE2 (1=SYSMIS).
END IF.
EXECUTE.
```

*Contradictive answers transformed in invalid/missing.

```
RECODE Q15 (3 thru 4=1) (ELSE=0) INTO OverEducation.
VARIABLE LABELS  OverEducation 'OE'.
EXECUTE.
```

*Create variable total relevant work experience.

```
COMPUTE Total_Rel_Exp=R_Exp_WEP + R_Exp.
EXECUTE.
```

*Skills, create average of two questions

```
COMPUTE S7=(S7A + S7B) / 2.
EXECUTE.

COMPUTE S9=(S9A + S9B) / 2.
EXECUTE.
```

*Reverse skills.

```
RECODE S2R (5=1) (4=2) (3=3) (2=4) (1=5) INTO S2.
VARIABLE LABELS  S2 'RevS2'.
EXECUTE.

RECODE S5R (5=1) (4=2) (3=3) (2=4) (1=5) INTO S5.
VARIABLE LABELS  S5 'RevS5'.
EXECUTE.

RECODE S6R (5=1) (4=2) (3=3) (2=4) (1=5) INTO S6.
VARIABLE LABELS  S6 'RevS6'.
EXECUTE.

RECODE S8R (5=1) (4=2) (3=3) (2=4) (1=5) INTO S8.
VARIABLE LABELS  S8 'RevS8'.
EXECUTE.

RECODE S13R (5=1) (4=2) (3=3) (2=4) (1=5) INTO S13.
VARIABLE LABELS  S13 'RevS13'.
EXECUTE.

RECODE S15R (5=1) (4=2) (3=3) (2=4) (1=5) INTO S15.
VARIABLE LABELS  S15 'RevS15'.
EXECUTE.

*Select only respondents that filled everything in

DATASET ACTIVATE DataSet1.
USE ALL.
COMPUTE filter_$=(Completed = 100).
VARIABLE LABELS filter_$ 'Completed = 100 (FILTER)'.
```

```
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.
* Encoding: UTF-8.
FACTOR
  /VARIABLES S1 S2 S3 S4 S5 S6 S7 S8 S9 S10 S11 S12 S13 S14 S15 S16 S17 S18 S19 S20 S21
  /MISSING LISTWISE
  /ANALYSIS S1 S2 S3 S4 S5 S6 S7 S8 S9 S10 S11 S12 S13 S14 S15 S16 S17 S18 S19 S20 S21
  /PRINT INITIAL EXTRACTION KMO ROTATION
  /FORMAT BLANK(.30)
  /CRITERIA MINEIGEN(1) ITERATE(25)
  /EXTRACTION PAF
  /CRITERIA ITERATE(25)
  /ROTATION PROMAX(4)
  /METHOD=CORRELATION.

FACTOR
  /VARIABLES S1 S2 S4 S5 S6 S7 S8 S9 S10 S11 S12 S13 S14 S15 S16 S17 S18 S19 S20 S21
  /MISSING LISTWISE
  /ANALYSIS S1 S2 S4 S5 S6 S7 S8 S9 S10 S11 S12 S13 S14 S15 S16 S17 S18 S19 S20 S21
  /PRINT INITIAL EXTRACTION ROTATION
  /FORMAT BLANK(.30)
  /CRITERIA FACTORS(6) ITERATE(25)
  /EXTRACTION PAF
  /CRITERIA ITERATE(50)
  /ROTATION PROMAX(6)
  /SAVE REG(ALL)
  /METHOD=CORRELATION.

FACTOR
  /VARIABLES S1 S2 S3 S4 S5 S6 S7 S8 S9 S10 S11 S12 S13 S14 S15 S16 S17 S18 S19 S20 S21
```

```
  /MISSING LISTWISE
  /ANALYSIS S1 S2 S3 S4 S5 S6 S7 S8 S9 S10 S11 S12 S13 S14 S15 S16 S17 S18 S19 S20 S21
  /PRINT INITIAL EXTRACTION ROTATION
  /FORMAT BLANK(.30)
  /CRITERIA FACTORS(7) ITERATE(25)
  /EXTRACTION PAF
  /CRITERIA ITERATE(50)
  /ROTATION PROMAX(4)
  /SAVE REG(ALL)
  /METHOD=CORRELATION.

DATASET ACTIVATE DataSet1.
RELIABILITY
  /VARIABLES=S1 S2 S3 S4 S5 S6 S7 S8 S9 S10 S11 S12 S13 S14 S15 S16 S17 S18 S19 S20 S21
  /SCALE('ALL VARIABLES') ALL
  /MODEL=ALPHA
  /STATISTICS=DESCRIPTIVE CORR
  /SUMMARY=TOTAL.
* Encoding: UTF-8.
*Create difference apparant education and genuine.
DO IF (Satisfaction <= 3).
RECODE A_OE (1=0).
END IF.
EXECUTE.

DO IF (Satisfaction >= 4).
RECODE G_OE (1=0).
END IF.
EXECUTE.

*
DATASET ACTIVATE DataSet1.
```

```
RECODE Edu_Direction (1=0) (2 thru 4=1) INTO DirectionD.
VARIABLE LABELS  DirectionD 'Dummy Direction'.
EXECUTE.

*Regressieanalyse (controleer Assumpties)
* Test Linear relationship

DATASET ACTIVATE DataSet1.
GRAPH
  /SCATTERPLOT(BIVAR)=Experience WITH Wage_ord
  /MISSING=LISTWISE.

DATASET ACTIVATE DataSet1.
GRAPH
  /SCATTERPLOT(BIVAR)=R_Exp WITH Wage_ord
  /MISSING=LISTWISE.

DATASET ACTIVATE DataSet1.
GRAPH
  /SCATTERPLOT(BIVAR)=Total_Rel_Exp WITH Wage_ord
  /MISSING=LISTWISE.

DATASET ACTIVATE DataSet1.
GRAPH
  /SCATTERPLOT(BIVAR)=Workhours WITH Wage_ord
  /MISSING=LISTWISE.

*Unemployed do not have wage / cannot be taken into account

USE ALL.
COMPUTE filter_$=(Unemployed = 0).
VARIABLE LABELS filter_$ 'Unemployed = 0 (FILTER)'.
```

```
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.

*Complete Regression with assumptions

REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS CI(95) R ANOVA COLLIN TOL
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT Wage_ord
  /METHOD=ENTER UnderEducation ReqEducation Experience Gender Exp2 DirectionD D_WEP
    Regio Selfemployed Judicia Edu_Level
  /PARTIALPLOT ALL
  /SCATTERPLOT=(*SRESID ,*ZPRED)
  /RESIDUALS DURBIN HISTOGRAM(ZRESID) NORMPROB(ZRESID)
  /CASEWISE PLOT(ZRESID) OUTLIERS(3)
  /SAVE PRED ZPRED ADJPRED MAHAL COOK LEVER RESID ZRESID SRESID SDBETA
SDFIT.

*Calculate Influential outliers

USE ALL.
COMPUTE filter_$=(COO_1 > 0.0377 | LEV_1 > 0.3207).
VARIABLE LABELS filter_$ 'COO_1 > 0.0377 | LEV_1 > 0.3207 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.
```

```
FREQUENCIES VARIABLES=Edu_Level Edu_Direction Judicia Experience Selfemployed Regio Workhours
    Wage_ord D_WEP UnderEducation OverEducation ReqEducation OE2 A_OE G_OE Total_Rel_Exp F1 F2 F3 F4 F5
    F6 F7
  /ORDER=ANALYSIS.

REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS CI(95) R ANOVA COLLIN TOL
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT Wage_ord
  /METHOD=ENTER UnderEducation ReqEducation Experience Exp2
  /PARTIALPLOT ALL
  /SCATTERPLOT=(*SRESID ,*ZPRED)
  /RESIDUALS DURBIN HISTOGRAM(ZRESID) NORMPROB(ZRESID)
  /CASEWISE PLOT(ZRESID) OUTLIERS(3).

REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS CI(95) R ANOVA COLLIN TOL
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT Wage_ord
  /METHOD=ENTER UnderEducation ReqEducation Experience Exp2 Gender Regio Judicia Selfemployed Edu_Level DirectionD D_WEP
    Total_Rel_Exp Workhours F1 F2 F3 F4 F5 F6 F7
  /PARTIALPLOT ALL
  /SCATTERPLOT=(*SRESID ,*ZPRED)
  /RESIDUALS DURBIN HISTOGRAM(ZRESID) NORMPROB(ZRESID)
  /CASEWISE PLOT(ZRESID) OUTLIERS(3).
```

```
* Logistic regression
DATASET ACTIVATE DataSet1.
USE ALL.
COMPUTE filter_$=(OverEducation = 1 | ReqEducation = 1).
VARIABLE LABELS filter_$ 'OverEducation = 1 | ReqEducation = 1 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.

LOGISTIC REGRESSION VARIABLES OverEducation
  /METHOD=ENTER Edu_Level Judicia Regio D_WEP DirectionD F1 F2 F3 F4 F5 F6
    Total_Rel_Exp
  /SAVE=PGROUP
  /CLASSPLOT
  /PRINT=GOODFIT CI(95)
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).

*Apparant overeducation + genuine
USE ALL.
COMPUTE filter_$=(A_OE = 1 | ReqEducation = 1).
VARIABLE LABELS filter_$ 'A_OE = 1 | ReqEducation = 1 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.

LOGISTIC REGRESSION VARIABLES A_OE
  /METHOD=ENTER Edu_Level Judicia Regio D_WEP DirectionD F1 F2 F3 F4 F5 F6
    Total_Rel_Exp
  /SAVE=PGROUP
  /CLASSPLOT
```

```
  /PRINT=GOODFIT CI(95)
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).

USE ALL.
COMPUTE filter_$=(G_OE = 1 | ReqEducation = 1).
VARIABLE LABELS filter_$ 'G_OE = 1 | ReqEducation = 1 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.

LOGISTIC REGRESSION VARIABLES G_OE
  /METHOD=ENTER Edu_Level Judicia Regio D_WEP DirectionD F1 F2 F3 F4 F5 F6
    Total_Rel_Exp
  /SAVE=PGROUP
  /CLASSPLOT
  /PRINT=GOODFIT CI(95)
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```