Thailand and The World Economy
P-ISSN: 2630-0931
E-ISSN: 2651-0529
Homepage : *https://www.tci-thaijo.org/index.php/TER*
Vol. 44, No.1, January - April 2026 Page [176-215]

The Effect of Vertical Educational Mismatch on Wages: Evidence from Indonesia's Waged Sector

*Dani Rahman Hakim**
Faculty of Economics and Business, Universitas Pamulang, Indonesia.
Iin Rosini
Faculty of Economics and Business, Universitas Pamulang, Indonesia.

Received 24 April 2024, Received in revised form 10 November 2024,
Accepted 15 November 2024, Available online 5 January 2026

Abstract

This study is the first to analyse the determinants and wage impacts of vertical educational mismatches in the wage sector in Indonesia. This study is also the first to examine the wage effect of vertical educational mismatch based on the fields of study, gender, and spatial conditions of the worker cohorts. Using the 2022 national labour force survey datasets, we calculate overeducation and undereducation using the job analysis approach. We found that almost half of the Indonesian workforce held jobs that did not correspond to their level of education. The determinants of vertical educational mismatch are identified based on individual, educational, and fields of study, job and employer, and spatial characteristics. Using the ORU[1] and VV[2] models, we found that the wage effect of vertical educational mismatch relied on gender, fields of study, and spatial conditions. In general, both overeducated and undereducated workers earn a wage premium. Overeducated workers in urban areas get higher returns than those in rural areas. However, if estimated based on the field of study cohort, there is a wage penalty for overeducated workers in at least seven fields. This study indicates that the mechanism of supply and demand for educated labour in each scientific field and industrial sector determines the wage impact of vertical educational mismatch.

Keywords*:* Educational Mismatch, Overeducation, Undereducation, Wages
JEL Classifications: I26; J24; J31

* **Corresponding author**: Email: danirahmanhak@gmail.com

1. Introduction

International Labour Organization (ILO) recorded that in 2021, only 47.6% of workers worldwide held jobs corresponding to their level of education. In contrast, 36.9% of workers are undereducated, and the rest, 15.5% are overeducated. The actual conditions could be more profound because this data only covers 130 countries. The high number of this vertical educational mismatch makes this topic still receive attention from scholars because it potentially reduces individual wage levels (Croce & Ghignoni, 2012; Diem, 2015; Serikbayeva & Abdulla, 2022) and economic growth (Abidin & Zakariya, 2018; Neycheva, 2020). Moreover, analysing the wage effect of vertical educational mismatch is one of the critical ways to evaluate education's external efficiency.

During the last two decades, many scholars have examined the determinants of vertical educational mismatch and its impact on wages. In estimating the wage effect of vertical educational mismatch, most scholars employ two approaches, which are ORU (overeducation, required education, and undereducation) from Duncan & Hoffman (1981) and the dummy variable approach from Verdugo & Verdugo (1989) (VV). The ORU approach developed Mincer's (1974) rate of return to education equation into three parts: overeducation (surplus schooling), required education, and undereducation (deficit schooling).

In the ORU model, if the level of education determines productivity, then the returns to overeducation and required education would be positive, while undereducation would be negative (Chung, 2001). It has been confirmed by many scholars, such as Kiker et al. (1997), Dolton & Vignoles (2000), Iriondo & Amaral (2016), and Clark et al. (2017). If, on the other hand, productivity was not determined by the level of education (but by other factors such as experiences, job training, or other human capital attributes), then the return to overeducation in the ORU model would be negative. Several studies, such as Verhaest & Omey (2012), Li & Miller (2015), Haddad & Habibi (2017), found a negative return to overeducation in the ORU model.

The ORU model compared the overeducated and undereducated workers with matched coworkers. Suppose matched workers get a higher rate of return than overeducated ones. In that case, it indicates that the increased years of education for overeducated workers would be inefficient (even if there were any positive returns). Meanwhile, the VV model compared the over- and undereducated workers to those matched in their jobs at the same education level. If education determines productivity and wages, the VV model predicted a positive overeducation coefficient and a negative undereducation coefficient. Overeducated workers would get higher wages than matched coworkers in the same job. However, their wages would be lower than those of other workers with the same level of education but who work according to their educational level (Kiker et al., 1997). Undereducated workers are predicted to get a wage premium that is higher than that of matched workers with the same level of education.

By employing the VV model, studies conducted by Bauer (2002), Iriondo & Amaral (2016), Park & Jang (2017), Johnes (2019), Carmichael et al. (2021), and Sun & Kim (2022) found a negative effect of overeducation on wages, while undereducation had a positive one. However, the overeducation coefficient on wages in the VV model was not always negative, nor was the coefficient of undereducation, which is not always positive. For instance, Cohn & Khan (1995) found a wage premium for overeducated workers and a wage penalty for undereducated workers in the VV model. It is strengthened by Bauer (2002), who found a wage premium for overeducated female workers if the VV model is generated by a fixed effect estimator. Carmichael et al. (2021)

also employed the VV model. They found that overeducated workers who work according to their field of education would get a wage premium. Meanwhile, undereducated workers experience a wage penalty.

The heterogeneity in the wage effect of undereducation and overeducation happens because of differences in theoretical perspectives. Referring to human capital theory (Becker, 1992), occupational mobility theory (Hout, 1984), and career mobility theory (Sicherman & Galor, 1990), overeducation is only a temporary phenomenon. These theories indicate that overeducated workers earn more while undereducated workers earn less. These theories produce the stepping stone hypothesis that the overeducated workers would finally get jobs or occupations according to their level of education. On the other hand, referring to the work assignment theory (Sattinger, 1993), wages are determined by the work assignment model. Workers with better work assignments would have higher wages even if undereducated.

This theoretical gap makes this topic still need further study, especially in countries with high levels of vertical educational mismatch, such as Indonesia. For instance, the ILO noted that 66.53 million Indonesian workers experienced a vertical educational mismatch in 2021. The composition was that 20.66 million workers were overeducated, and 45.87 million were undereducated. It is approximately 49.51% compared to Indonesia's total workforce in 2021. Despite the vertical educational mismatch in Indonesia being quite apprehensive, studies that analyse the vertical educational mismatch in this country are still rarely conducted.

Several scholars who have discussed the vertical educational mismatch in Indonesia are Mugijayani (2020), Sitorus & Wicaksono (2020), and Wulandari & Damayanti (2021). However, their studies have several shortcomings. For instance, Mugijayani (2020) employed the Indonesian Family Life Survey (IFLS) 2000 and 2014, a relatively outdated dataset. Likewise, Sitorus & Wicaksono (2020) discussed the impact of vertical educational mismatch on wages less comprehensively. Meanwhile, Wulandari & Damayanti (2021) did not discuss the determinants of vertical educational mismatch.

The deficiencies of the previous study were the primary motivation for conducting this study. We examine the determinants of vertical educational mismatch and its impact on Indonesian labour wages. The lack of studies on vertical education mismatch in Indonesia is perhaps due to the complexity of educational mismatch and limited access to employment datasets. A relatively large employment dataset is required to examine vertical educational mismatch more comprehensively. Fortunately, we were given access to Indonesia's National Labour Survey 2022 datasets conducted by the Central Statistics Agency of Indonesia. Therefore, this study could become a more comprehensive and detailed literature discussion of vertical education mismatch in the context of the Indonesian waged sector. Hopefully, this study could provide productive input for the Indonesian government in formulating effective policies to minimise this vertical educational mismatch.

2. Literature Review

Several scholars, such as Robert (2014), Pholphirul (2017), and Li et al. (2018), used the term vertical education mismatch to describe workforce mismatch based on education level. It is then divided into overeducation and undereducation (McGuinness et al., 2018; Wu & Wang, 2018). In this context, overeducation is a surplus of years of education, while undereducation is a deficit of years of education (Rumberger, 1981). A worker is overeducated if he or she gets a job whose educational requirements are below his or her educational qualifications. On the other hand, a worker is undereducated if their educational level is lower than the qualifications required for the job.

The minimum education required for a job is necessary to calculate overeducation and undereducation. If the required education has been determined, then overeducation and undereducation can be calculated. Thus, vertical educational mismatch is measured to determine the required education. Munsech (2019) states that measuring overeducation and undereducation is divided into two approaches: objective and subjective. The subjective approach asks respondents about the minimum education required to get a job. Meanwhile, the objective approach is carried out in 2 forms: normative with job analysis (JA[3]) and statistical.

The JA normative approach determines the required education by identifying the International Standard Classification of Occupations (ISCO[4]) job code classification or other relevant classifications. Statistical approaches (often called realised matched (RM[5])), on the other hand, determine the required match based on the mean value, range, or mode of the distribution of the workforce's education level in a particular occupation. For example, Verdugo & Verdugo (1989) used the mean, whereas Kiker et al. (1997) used the mode. Verdugo & Verdugo (1989) determine that workers are overeducated if their education level is higher than one standard deviation, whereas if it is less than one standard deviation, they are undereducated. However, if the worker's years of education are -/+ in the range of 1, then the worker is matched.

All vertical educational mismatch measurement approaches are debated and criticised (Verhaest & Omey, 2006). However, according to Hartog (2000), the RM approach inaccurately reveals technology needs as work requirements. Therefore, Hartog (2000) explains that the JA approach is conceptually considered better. On the other hand, Verhaest & Omey (2006) explain that although there would be differences in the magnitude of the vertical educational mismatch coefficient on wages, the results remain robust from various measurements. It means that the effect of vertical educational mismatch on wages would result in a similar coefficient of statistical power.

2.1. Determinants of vertical educational mismatch

Many scholars have attempted to analyse the determinants of vertical educational mismatch. They employed several theories such as human capital theory (Becker, 1992), differential overqualification theory (Frank, 1978), career mobility theory (CMT[6]) (Sicherman and Galor, 1990), job competition theory (JCT[7]) (Thurow, 1975), matching theory (Jovanovic, 1979), job market signalling theory (Spence, 1973), and spatial mismatch theory (Kain, 1992). The human capital theory and the differential overqualification theory could explain the determinants of vertical educational mismatch based on the perspective of individual characteristics.

Human capital theory indicates that vertical educational mismatch is determined by the attributes of an individual's human capital, such as the level of education, health, experience, and skills from training. Meanwhile, the differential overqualification theory explains that individual characteristics such as gender and marital status determine vertical educational mismatch. According to this theory, a married woman is likelier to be overeducated (McGuinness et al., 2018). Some studies have proved these theories. For instance, Verhaest & Omey (2010), Leuven & Oosterbeek (2011), Devillanova (2013), Senkrua (2015), and Sitorus & Wicaksono (2020) found that individual characteristics affect vertical educational mismatch. Nevertheless, individual characteristics cannot always determine vertical educational mismatch. Several other studies, such as Piracha et al. (2012) and Caroleo & Pastore (2018), found no effect of gender and marital status on overeducation or undereducation.

On the other hand, CMT (Sicherman & Galor, 1990) and matching theory (Jovanovic, 1979) could describe the determinants of vertical educational mismatch based on the type of job and employer characteristics. CMT postulates that workers with higher education tend to move to higher jobs (Sicherman, 1990). One of the hypotheses of CMT is that vertical mismatch is a stepping stone that is part of career mobility (Blázquez & Budría, 2012). Therefore, workers who do not receive appropriate promotions tend to leave their old jobs even if their education level is high (Sicherman & Galor, 1990). Thus, CMT assumes workers in companies with larger sizes and better career systems have a relatively lower probability of experiencing vertical educational mismatch.

The matching theory reinforced the CMT. The matching theory indicates that employers can conduct job recruitment based on the suitability of qualifications. The process of matching these qualifications is carried out in recruitment. Workers in larger companies with a better recruitment system would have a lower probability of experiencing vertical educational mismatch. The relevance of CMT and matching theory is proven by several studies, including Karakaya et al. (2007), Belfield (2010), Zakariya & Noor (2014), and Ege & Erdil (2023). Their studies found that several job and employer characteristics, such as company size, working hours system, in-job training, recruitment system, and company credibility, can be determinants of overeducation and undereducation.

The characteristics of educational institutions are also considered to be a determinant of vertical educational mismatch. Job market signalling theory from Spence (1973) suggests that employers recruit workers based on signals given by prospective workers. These signals can be in the form of characteristics of prospective workers, such as level of education, skills, and reputation of educational institutions. Prospective workers from reputable educational institutions (for instance, from the ten ranked universities) are considered to have better skills so that they would receive higher wage offers. On that basis, the type of educational institutions would determine the vertical mismatch of education (Chevalier & Lindley, 2009). Several other determinants of educational vertical mismatch that originate from the educational institution's characteristics include the type of vocational education (McGuinness et al., 2018) and field of study (Carroll & Tani, 2015; Zheng et al., 2021).

In addition, spatial characteristics could also affect vertical educational mismatch. It draws from JCT (Thurow, 1975) and spatial mismatch theory (Kain, 1992). JCT indicates that educational mismatch occurs due to rigid demand for educated labour caused by spatial and technological reasons. Companies would choose operational locations in areas with adequate labour supply. On the other hand, workers in suburban or rural areas are relatively unable to access jobs appropriate to their educational level. As a result, these workers would migrate to areas with higher demand for labour, resulting in increasingly tight job competition. It triggers a high level of vertical educational mismatch.

The JCT theory is supported by spatial mismatch theory (Kain, 1992), which states that workers in suburban or rural areas would have a higher possibility of experiencing vertical education mismatch due to limited labour demand. Several studies, including Quinn & Rubb (2006), Croce & Ghignoni (2012), and Devillanova (2013) found that spatial factors are determinants of vertical educational mismatch.

2.2. The effect of vertical educational mismatch on wages

Many scholars employ two models to examine the effect of vertical educational mismatch on wages: ORU and VV (stands for Verdugo & Verdugo (1989)). In the ORU model, Duncan & Hoffman (1981) decomposed years of educational attainment (S^{α}) from

Mincer's (1974) model with S^0, S^r, dan S^u. S^0 is surplus schooling or overeducation, S^r is school required or the education needed to get a job, while S^u is deficit schooling or undereducation. The standard form of the ORU model is as follows:

$$\mathrm{LnW_i} = \mathrm{X_i}\delta + \Upsilon_1 \mathrm{S_i^r} + \Upsilon_2 \mathrm{S_i^0} + \Upsilon_3 \mathrm{S_i^u} + \mathrm{u_i} \quad (1)$$

LnWi is the natural logarithm of individual wages (for example, per hour, month, or year), while X_i is a vector that measures worker characteristics. The S^r variable is the years of education required to get a job, while S^o and S^u are overeducation and undereducation.

The ORU model is based on human capital theory, which assumes that the higher an individual's education or training would return the higher productivity and wages. In the ORU model, returns to overeducation and required education are predicted to be positive, while returns to undereducation are negative. If there is a positive return, overeducation still has economic value as an additional wage for individuals. However, this is a form of underutilisation borne by the employer (Duncan & Hoffman, 1981; Hartog & Oosterbeek, 1988). Employers give higher wages to overeducated workers even though the work tasks carried out by these workers do not match their level of education.

By using the ORU model, several studies, such as Chung (2001), Bauer (2002), Groeneveld & Hartog (2004), Korpi & Tåhlin (2009), Zakariya (2014), Grunau & Pecoraro (2017), and Clark et al. (2017), found positive returns to overeducation and negative returns to undereducation. The positive return to overeducation shows that an increase in 1 year of education, even though excessive, would still increase wages, implying that additional education for overeducated workers can boost their productivity and be appreciated by employers. As explained by job signalling theory, employers consider that the level of education is a signal of a worker's productivity.

However, a year increase in education by overeducated workers does not always produce a positive return. Studies by Verhaest & Omey (2012), Li & Miller (2015), and Haddad & Habibi (2017) found a negative return to overeducation in the ORU model. If overeducation produces negative returns, the additional level of education attained is viewed as an inefficient human capital investment. The increase in years of education for overeducated workers is not rewarded in the form of wages by employers. In this condition, productivity is not determined by the level of education but by other factors such as experience, in-job training, soft skills, and others. As work assignment theory states, wages are determined more by work assignments, not just education level. It is strengthened by Groot (1996), who states that the cause of the negative return on overeducation is the lack of productivity of overeducated workers.

A model that has also received much adoption from scholars in studying the effect of vertical educational mismatch on wages is the VV model. This model modifies the ORU model by replacing S^0 and S^u with dummy variables overeducation (OE) and undereducation (UE). In contrast, S^r is replaced by the level of education the individual attains (Educ). The main aim of this model is to examine the wage penalty and wage premium more clearly as follows:

$$\mathrm{LnW_i} = \mathrm{X_i}\delta + \beta_1 \mathrm{Educ_i} + \beta_2 \mathrm{OE_i} + \beta_3 \mathrm{UE_i} + \mathrm{u_i} \quad (2)$$

LnW is the natural logarithm value of individual wages, while X_i vectors control individual heterogeneity, such as experience, age, in-job training, and others. $Educ_i$ is the year of school completed by the individual, while OE_i and UE_i are dummy variables for overeducation and undereducation.

In the VV model, the wages earned by overeducated and undereducated workers are compared with other workers with the same characteristics (including their level of education) but who work in other jobs with the required years of education that they have. For example, 3 Diploma 3 graduates have 15 years of education. The first graduate works as a technician (matched), the second graduate works as a service business worker (overeducated), and the third graduate works as a professional (undereducated). The wage level for overeducated graduates is predicted to be lower than that of matched graduates, while the wage for undereducated graduates will be higher. It means that the coefficient produced by undereducated workers will be positive (wage premium), while the coefficient for overeducated workers will be negative (wage penalty).

This VV model is relatively in line with the work assignment theory (Sattinger, 1993), which states that the level of education and work assignments determine wages. The workers would earn more if they had more assignments from their employers. Based on this assumption, undereducated workers are predicted to get higher wages than matched workers at the same level of education. It also aligns with Nash bargaining, which predicts wage penalties for overeducation and wage rewards for undereducation (Sattinger & Hartog, 2013). By employing the VV model, several scholars, such as Bauer (2002), Cutillo & Pietro (2006), Diem & Wolter (2014), Iriondo & Amaral (2016), Park & Jang (2017), Johnes (2019), Schweri et al. (2020), Carmichael et al. (2021), and Sun & Kim (2022), found a wage penalty for overeducated workers and a wage premium for undereducated workers.

If the overeducation coefficient in the VV model is negative, then overeducated workers receive a lower wage than matched workers at the same level of education. On the other hand, if the overeducation coefficient is positive, then overeducated workers get higher wages than matched workers with the same level of education. This condition occurs if the employer prioritises aspects of education level rather than productivity. In other words, the wage level given by employers is based on the level of education achieved, not on work productivity.

The positive coefficient of overeducation on wages in the VV model can also occur because there are quite striking differences in wage levels between employment sectors and the size of the employing company. For instance, a bachelor of accounting who works as an internal junior accountant (overeducated) in a state-owned company would probably earn a higher wage than an accounting graduate who works as an accountant (matched) in a start-up public accounting firm. Several studies that found overeducated workers received a wage premium using the VV model include Cohn & Khan (1995), Bauer (2002), Bedir (2014), and Carmichael et al. (2021).

3. Method

This study employs data from Indonesia's Labour Force Survey 2022. It is Indonesia's largest survey of labour market conditions. In 2022, Indonesia's employment and economic conditions recovered from the impact of the COVID-19 pandemic. Evidently, Indonesia's economic growth was 5.6% that year. Therefore, this employment data is not significantly affected by the pandemic. This survey was conducted in 38 provinces in Indonesia with a total population of 209,420,383 people. This survey's population is all working-age Indonesian citizens (over 15 years). This survey took 752,688 respondents as the total sample. The sampling method in this survey has two stages, with one phase of stratified sampling. However, we do not include the entire sample when estimating vertical educational mismatch.

In estimating the determinants of vertical educational mismatch and its impact on wages, we excluded government employees because career paths in the government bureaucracy are relatively more procedural and prioritise aspects of work experience. This study excluded samples with self-employed status. We only used samples with labour status (employees or labourers) in the waged sector. We also excluded workers who did not disclose information regarding their educational level, military occupational groups (ISCO code 0), and elementary occupations (ISCO code 9). The final sample used in this study was 76,747, with the following distribution:

Table 1: Indonesia's Private Waged Sector Workers 2022

ISCO Code	Occupations	Sample	%	Population	%
1	Managers	2,039	2.66	777,767	2.67
2	Professionals	8,371	10.91	3,159,838	10.86
3	Technicians and Associate Professionals	5,328	6.94	2,225,605	7.65
4	Clerical Support Workers	8,534	11.12	3,399,881	11.69
5	Services and Sales Workers	22,041	28.72	8,210,860	28.22
6	Skilled Agricultural, Forestry and Fishery Workers	3,542	4.62	808,467	2.78
7	Craft and Related Trades Workers	14,595	19.02	5,617,289	19.31
8	Plant and Machine Operators and Assemblers	12,297	16.02	4,896,390	16.83
	TOTAL	**76,747**	**100**	**29,096,097**	**100**

Notes: *The % value in the sample is obtained by dividing the number of samples for each occupational group by the total sample and then multiplying by 100. For example, 2.66% is obtained by (2,039/76,747*100). The same method is used to calculate the % value in the population. This process also validates the distribution equivalence to generalise the sample to the population.*

Source: Calculated by authors from the raw datasets of the Indonesia's Labour Force Survey

From the sample size in Table 1, we estimate overeducation and undereducation using the JA normative method. In the JA approach, this study determines the required education for managers and professional jobs of 16 years (equivalent to a bachelor's degree). For the technicians and associate professionals category, it is 15 years (equivalent to a third diploma), while for the other categories, it is 12 years (equivalent to high school), respectively. This categorisation is based on matching the 2008 ISCO code and the 2013 International Standard Classification of Education (ISCED[8]) 2013. However, according to these standards, managers and professionals will not experience overeducation because the skill levels for these two occupations are levels 3 and 4. Nevertheless, the education required to become a manager or professional also relates to the company institutions.

Managers of large companies, such as state-owned enterprises (SOE[9]), require higher skills than retail managers. On the other hand, this study's dataset does not provide specific information regarding the type of manager's company, so SOE managers cannot be differentiated from retail store managers. Based on this, if the skill levels for managers and professionals are all set at levels 3 and 4, there is potential for bias in measuring vertical mismatch and its effect on wages. It is because the type of company also determines a manager's wages. For example, the wages of a manager in an SOE and a retail store manager can be very different. Based on these facts, the required match for manager and professional occupations is equivalent to a bachelor's degree (16 years of school).

We also measured vertical educational mismatch using the RM method, which Verdugo & Verdugo (1989) originated to compare the determinants of vertical educational mismatch. In this RM approach, workers whose education level is higher than one standard deviation from the average of the labour group in the first digit ISCO code are declared to be overeducated. In contrast, they are undereducated if their education is less than one standard deviation. However, we employ a size of ½ standard deviation because this study's sample size is relatively large. We subdivided the workforce groups for each first-digit ISCO code into six age groups. These age groups include young people (14-24 years), young workers (25-34), middle-aged (35-44), pre-retirement age (45-54), retirement group (55-64), and elderly (above 65). We include the 55 and above age group because even though the normal retirement age in Indonesia is 64, most are still working. Primarily those working in the middle companies or those not working in government agencies. Therefore, we assume that including this age group in the analysis is still useful.

Furthermore, this study examines the determinants of vertical education using a logistic regression (Logit) estimator. The equation tested to examine the determinants of overeducation is as follows:

$$L_i\mathrm{DOVER} = \ln\left(\frac{P_i}{1-P_i}\right) = \alpha + \beta_k X_{ki} + u_i \tag{3}$$

L_iDOVER[10] is an overeducation dummy that is measured categorically (filled with 1 for overeducation and filled with 0 if not) in the form of a log odds ratio. Meanwhile, Xk is vectors in the form of individual characteristics, educational institution characteristics, job and employer characteristics, and spatial characteristics. DOVER comes from the assumption $S^\alpha > S^r = S^o$ (overeducation), $S^\alpha = S^r$ = matched, dan $S^\alpha < S^r = S^u$ (undereducation). In this case, S^α is school attained or years of school achieved/completed, Sr is the school required or years needed to get a job, S^o is overeducation or surplus schooling, and S^u is undereducation or deficit schooling.

As for examining the determinants of undereducation, the equation that will be tested is as follows:

$$L_i\mathrm{DUNDER} = \ln\left(\frac{P_i}{1-P_i}\right) = \alpha + \beta_k X_{ki} + u_i \tag{4}$$

LiDUNDER[11] is the Logit value or log odds ratio of undereducation. The α value is a constant, while Xk is vectors in the form of individual characteristics, educational institution characteristics, job and employer characteristics, and spatial characteristics.

Furthermore, this study employs ORU and VV models to examine the wage effect of vertical educational mismatch. The ORU model that would be estimated is as follows:

$$\mathrm{LnW_i} = \beta_0 + \beta_1 \mathrm{YOVER_i} + \beta_2 \mathrm{YREQ_i} + \beta_3 \mathrm{YUNDER_i} + \beta_k X_{ki} + \varepsilon_i \tag{5}$$

LnW is the per-hour wage in the form of a natural logarithm. At the same time, YOVER[12], YREQ[13], and YUNDER[14] are the number of years of surplus education (overeducation), years required, and years of deficit education (undereducation). The YOVER value is YEDUC – YREQ if YEDUC > YREQ. In this case, YEDUC[15] is the number of years of education attainment of individual i, while the YUNDER value is YREQ – YEDUC if YEDUC < YREQ. YOVER, YREQ, and YUNDER coefficients will show the return to overeducation, required education, and undereducation. Meanwhile, βk is the coefficient of Xk, a vector to control individual heterogeneity. The variables Xk are experience and age, while u is the error term. Experience and age in our dataset are two unrelated variables. The experience here is the number of years a worker has worked at his or her current job and company. Therefore, older workers do not always have more remarkable experiences than young ones.

Meanwhile, our VV model is as follows:

$$\mathrm{LnW_i} = \beta_0 + \beta_1 \mathrm{EDUC_i} + \beta_2 \mathrm{DOVER_i} + \beta_3 \mathrm{DUNDER_i} + \beta_k \mathrm{X_{ki}} + \mathrm{u_i} \quad (6)$$

LnW is the natural logarithm of wage per hour i. EDUC is the number of years of school attained by individual i, while DOVER and DUNDER are overeducation and undereducation, which are measured categorically (dummy variables). Meanwhile, Xk are age and experience variables to control individual heterogeneities, while u is the error term.

This study employed the ordinary least squares (OLS[16]) estimator to estimate both models. We prefer OLS because our sample does not suffer from selection bias. It is because we have excluded workers who are currently unemployed. In addition, the wage variable in our study is the natural logarithm of the hourly wage, so there is no potential for selection bias for part-time workers. Several scholars who also used OLS with the wage per hour as a dependent variable are Kiker et al. (1997), Rubb (2006), Tsai (2010), Clark et al. (2017), Wen & Maani (2018) and others.

This study detects multicollinearity by correlating all explanatory variables. Furthermore, we employ a histogram to test the normal distribution assumption. The Y axis in the histogram is the normal k density, while the X axis is the residual from the OLS regression results. If most residual data is within the normal density line, then the error term data is distributed normally. This study employs histograms because our samples are relatively large, so statistical methods such as Jarque-Berra, Shapiro-Wilk, and others are ineffective. This study uses a scatterplot with the Y axis as the residual and the X axis as the fitted value or linear prediction from the regression analysis results to detect heteroscedasticity. If the fitted value data is distributed constantly, the data is homoscedastic. If, on the other hand, the data indicated heteroscedasticity, then we overcome this problem by using cluster-robust standard errors.

4. Results and Discussion

4.1. The incidences of vertical educational mismatch in Indonesia

Before estimating the regression model, this study attempts to capture the incidence of vertical educational mismatch using the JA approach and describe the average wage obtained by workers. The results are as follows:

Table 2: Descriptive Statistics of Wage by Vertical Educational Mismatch

	Under	Match	Over	Total
Male's wage per hour				
Mean	16,079.05	18,551.35	31,876.15	18,851.79
Median	13,000.00	14,583.30	22,058.80	1,4285.70
Std.Dev	20,197.01	22,920.92	56,702.36	27,578.69
Min	156.25	297.62	1,111.11	156.25
Max	1,100,000.00	1,300,000.00	2,800,000.00	2,800,000.00
Obs	20,589	25,180	4,964	50,733
%	40.58	49.63	9.78	100

	Under	Match	Over	Total
Female's wage per hour				
Mean	11,242.86	14,229.74	23,869.90	14,603.41
Median	8,035.71	10,416.70	18,229.20	10,416.70
Std.Dev	13,823.76	21,640.56	26,161.62	20,582.88
Min	306.12	260.42	441.18	260.42
Max	375,000.00	1,600,000.00	625,000.00	1,600,000.00
Obs	8,513	13,855	3,646	26,014
%	32.72	53.26	14.02	100

Notes: The unit of value used for wage per hour is the Indonesian Rupiah (IDR)
Source: Calculated by authors from the raw datasets of Indonesia's Labour Force Survey 2022

Assuming 1 USD is 15,000 IDR, the average hourly wage for male workers in the Indonesian wage sector is only 1.25 USD. If this value is multiplied by the regular working hours per week of 40 hours and then multiplied by four, the average male worker in Indonesia earns 201 USD per month. Meanwhile, the average wage for female workers is much lower, only 0.973 USD per hour or 155.7 USD per month. Moreover, a standard deviation value higher than the average indicates a wage level gap. It is partly due to the gap between rural and urban areas in 38 provinces in Indonesia.

Table 2 shows that only 50.86% of workers in Indonesia's waged sector work according to their educational level. If classified based on gender, only 49.63% of the male workforce and 53.26% of the female workforce were well-matched. Male workers experience overeducation as much as 9.78%, while for females, it is 14.02%. From this data, 11.22% of the workforce are overeducated, and 37.92% are undereducated. An overview of the incidence of educational mismatch by the first digit of the ISCED code is as follows:

Table 3: The Incidences of Vertical Educational Mismatch in Indonesia's Waged Sector (JA Approach)

ISCO Code	Occupations	Under (%)	Match (%)	Over (%)
1	Managers	52.82	41.44	5.74
2	Professionals	38.23	54.46	7.31
3	Technicians and Associate Professionals	64.23	11.88	23.89
4	Clerical Support Workers	5.02	53.50	41.48
5	Services and Sales Workers	26.56	63.87	9.57
6	Skilled Agricultural, Forestry and Fishery Workers	75.24	23.01	1.75
7	Craft and Related Trades Workers	50.29	46.54	3.17
8	Plant and Machine Operators and Assemblers	41.60	54.87	3.53
	TOTAL	**37.92**	**50.86**	**11.22**

Source: Calculated by authors from the raw datasets of Indonesia's Labour Force Survey 2022

Table 3 shows that almost half of the waged sector workforce in Indonesia do not work according to their education level. Several occupations with a high level of vertical educational mismatch are technicians and associate professionals, skilled agricultural workers, forestry and fishery workers, and craft and related trades workers. If estimated by other methods, such as the RM method, the percentage of vertical mismatch shown in Table 3 may differ. However, scholars have disagreed on which mismatch estimation method is the most accurate. If we refer to Hartog (2000), the JA approach better estimates educational mismatches caused by technological developments.

Most workers in the technicians and associate professionals groups are undereducated because the required education for this occupation is 15 years (equivalent to a diploma 3). Meanwhile, as many as 79.54% of workers working in the waged sector of Indonesia have an education below a diploma 3. In Indonesia, many technicians are graduates of vocational high schools, equivalent to 12 years of education. Likewise, with skilled agricultural workers, as many as 75.2% have an education below high school (12 years), so most of these workers are undereducated. Ideally, skilled agricultural workers have adequate levels of education. However, because there are still few agricultural majors in Indonesia (one of the reasons is because there are few enthusiasts), these skilled agricultural workers are filled with poorly educated workers. However, they usually get guidance and additional training from government agricultural instructors.

The largest percentage of overeducation is in the clerical support workers group because even though the workload is 'only' equivalent to a high school, many graduates occupy this occupation. This condition is partly caused by the increasing standard of education required for the clerical support worker's occupation. In larger companies, for example, fresh graduates with bachelor's degrees are eyeing the clerical and support worker positions. Although initially overeducated, highly educated workers who occupy clerical and support worker positions believe that the income offered will be higher. Moreover, employers offer a career ladder to this occupation.

4.2. Determinants of vertical educational mismatch in Indonesia's waged sector

This study employs JA and RM measurement methods to examine the determinants of overeducation and undereducation. The two educational vertical mismatch variables are categorical (1 and 0). This study uses two measures of goodness of fit in each logistic regression model: LR Chi^2 and Pseudo R^2. LR Chi^2 is used for simultaneous model testing (Gujarati, 2015). The null hypothesis for the LR Chi^2 statistic is that the explanatory variables cannot explain the dependent variable. Therefore, if the p-value is less than 0.05, the null hypothesis is rejected so that the explanatory variables can explain the dependent variable. Meanwhile, if the Pseudo R^2 value is in the range of 0.2 to 0.4, then the model can be declared to have goodness of fit (GOF[17]), whereas if it is more than that, it has excellent fit (McFadden, 2021).

We classify the determinants of vertical educational mismatch based on two criteria: secondary and tertiary education. The aim is to avoid education attainment bias because if it is not classified, workers with secondary education have a higher probability of being undereducated than those with tertiary education. The odds ratio from the determinants of undereducation in Indonesia's waged sector are as follows:

Table 4: Determinants of Undereducation

	DUNDER_JA		**DUNDER_RM**	
	Secondary Educ	**Tertiary Educ**	**Secondary Educ**	**Tertiary Educ**
LR Chi^2 (Prob)	6995	1608	1090	327
	(0.000)	(0.000)	(0.000)	(0.000)
Pseudo R^2	0.226	0.253	0.021	0.016
Obs	37,664	13,218	37,664	14,862
Constant	0.154***	0.001***	0.439***	0.358***
	(0.039)	(0.001)	(0.064)	(0.068)
Individual Characteristics				
Female (male as base)	0.929	1.222**	0.993	0.982
	(0.044)	(0.113)	(0.028)	(0.037)
Married (not married/divorced as base)	1.058	1.032	1.002	1.016
	(0.046)	(0.111)	(0.028)	(0.045)
Having Children (none as a base)	0.941	0.992	0.98	0.952
	(0.043)	(0.101)	(0.028)	(0.041)

	DUNDER_JA		DUNDER_RM	
	Secondary Educ	**Tertiary Educ**	**Secondary Educ**	**Tertiary Educ**
Age Group (15-24 as a base)				
25 – 34 years	1.151**	0.904	2.238***	2.43***
	(0.062)	(0.125)	(0.07)	(0.157)
35 – 44 years	1.216***	1.033	1.744***	1.814***
	(0.074)	(0.172)	(0.064)	(0.135)
45 – 54 years	1.244**	1.35	1.912***	2.012***
	(0.086)	(0.259)	(0.083)	(0.172)
55 – 64 years	1.584***	1.793**	2.128***	2.31***
	(0.16)	(0.45)	(0.147)	(0.251)
more than 64	2.864***	1.578	2.623***	4.785***
	(0.647)	(0.839)	(0.46)	(1.23)
Experience Group (1-5 years as a base)				
6-10 years	1.298***	1.423**	1.002	0.961
	(0.065)	(0.16)	(0.032)	(0.046)
11-15 years	1.261***	1.875***	0.978	1.022
	(0.077)	(0.26)	(0.039)	(0.061)
>15 years	1.601***	1.796***	1.083**	1.059
	(0.095)	(0.271)	(0.044)	(0.067)
Training (No as a base)	1.103**	0.835**	0.945**	1.047
	(0.04)	(0.069)	(0.023)	(0.037)
Education Characteristics				
Education Institution Type (Public as a base)				
Private school	0.913**	1.076	1.007	0.972
	(0.034)	(0.097)	(0.024)	(0.035)
Agency School	0.837	-	1.048	1.417
	(0.535)		(0.401)	(0.925)
Others School	0.545	-	1.171	-
	(0.236)		(0.274)	
ISCED 2 Digits (ISCED Code 011 as a base)				
ISCED Code 018	-	-	-	0.979
				(0.31)
ISCED Code 021	-	111.876***	-	1.141
		(44.178)		(0.172)
ISCED Code 022	1.484	9.493***	0.786	0.966
	(0.379)	(3.831)	(0.119)	(0.091)
ISCED Code 023	1.236	26.83***	1.018	0.686**
	(0.329)	(10.838)	(0.152)	(0.09)
ISCED Code 028	-	-	-	0.86
				(0.142)
ISCED Code 030	-	-	-	0.979
				(0.135)
ISCED Code 031	0.974	5.585***	0.846	0.929
	(0.202)	(2.445)	(0.091)	(0.069)
ISCED Code 032	-	1125.941***	-	0.904
		(1147.784)		(0.839)
ISCED Code 038	-	209.858***	-	1.24
		(97.884)		(0.348)
ISCED Code 041	0.905	33.448***	0.855	0.934
	(0.193)	(11.457)	(0.094)	(0.062)
ISCED Code 042	-	-	-	0.746**
				(0.074)

	DUNDER_JA		DUNDER_RM	
	Secondary Educ	**Tertiary Educ**	**Secondary Educ**	**Tertiary Educ**
ISCED Code 048	-	109.984*** (144.481)	-	0.312 (0.351)
ISCED Code 050	1.285 (0.268)	-	0.847 (0.092)	-
ISCED Code 051	-	2.287 (2.411)	-	0.972 (0.143)
ISCED Code 052	-	9.076** (9.659)	-	0.752 (0.183)
ISCED Code 053	2.579** (1.091)	59.734*** (27.511)	0.585 (0.207)	1.136 (0.203)
ISCED Code 054	-	-	-	1.007 (0.151)
ISCED Code 061	1.519* (0.336)	38.095*** (13.47)	0.892 (0.105)	0.877* (0.068)
ISCED Code 068	-	242.982*** (225.595)	-	1.364 (0.983)
ISCED Code 070	-	522.109*** (527.755)	-	0.687 (0.632)
ISCED Code 071	1.186 (0.25)	86.017*** (31.11)	0.866 (0.095)	0.891 (0.09)
ISCED Code 072	0.891 (0.386)	-	0.601** (0.154)	0.919 (0.129)
ISCED Code 073	1.703** (0.419)	28.659*** (12.014)	0.83 (0.122)	0.717** (0.079)
ISCED Code 078	1.788** (0.401)	215.758*** (93.534)	0.874 (0.111)	0.8 (0.194)
ISCED Code 079	1.489 (0.362)	61.031*** (25.842)	0.827 (0.116)	0.882 (0.128)
ISCED Code 080	-	-	-	-
ISCED Code 081	1.469 (0.366)	21.362*** (9.001)	0.987 (0.141)	0.921 (0.097)
ISCED Code 082	2.5 (2.935)	117.306*** (70.805)	0.398 (0.453)	0.691 (0.24)
ISCED Code 083	1.82** (0.522)	132.001*** (67.487)	0.779 (0.139)	1.032 (0.266)
ISCED Code 084	5.385** (4.569)	-	0.491 (0.349)	0.653 (0.573)
ISCED Code 088	-	-	-	-
ISCED Code 090	-	48.429*** (41.007)	-	0.735 (0.339)
ISCED Code 091	4.625*** (1.194)	134.246*** (46.727)	0.843 (0.141)	0.948 (0.09)
ISCED Code 098	-	78.761*** (97.638)	-	0.825 (0.84)
ISCED Code 101	0.88 (0.22)	71.858*** (32.84)	0.759** (0.096)	1.057 (0.192)
ISCED Code 102	-	176.048*** (178.086)	-	1.098 (0.862)
ISCED Code 103	-	-	-	1.676 (2.428)
ISCED Code 104	-	106.56*** (57.776)	-	0.726 (0.188)

	DUNDER_JA		DUNDER_RM	
	Secondary Educ	Tertiary Educ	Secondary Educ	Tertiary Educ
Job and Employer Characteristics				
Industry Sectors (ISIC Code A as a base)				
ISIC Code B	1.705***	1.343	1.013	0.865
	(0.195)	(0.698)	(0.076)	(0.164)
ISIC Code C	1.013	1.167	0.992	1.139
	(0.102)	(0.534)	(0.059)	(0.183)
ISIC Code D	5.302***	1.409	0.886	0.923
	(0.737)	(0.799)	(0.103)	(0.205)
ISIC Code E	2.492***	-	1.115	1.118
	(0.527)		(0.192)	(0.368)
ISIC Code F	3.079***	1.114	0.98	1.35*
	(0.344)	(0.564)	(0.073)	(0.237)
ISIC Code G	0.588***	1.001	0.984	0.987
	(0.064)	(0.456)	(0.059)	(0.157)
ISIC Code H	0.891	0.79	1.111	1.056
	(0.109)	(0.411)	(0.077)	(0.186)
ISIC Code I	0.906	1.615	1.069	1.1 (0.2)
	(0.117)	(0.788)	(0.074)	
ISIC Code J	6.604***	3.81**	1.03	1.205
	(0.835)	(1.773)	(0.1)	(0.213)
ISIC Code K	2.119***	1.168	0.951	1.086
	(0.241)	(0.534)	(0.072)	(0.173)
ISIC Code L	2.86***	0.636	1.285	1.084
	(0.589)	(0.538)	(0.207)	(0.267)
ISIC Code M and N	2.57***	1.404	0.992	1.079
	(0.301)	(0.675)	(0.079)	(0.183)
ISIC Code O	3.495***	2.027	1.055	0.931
	(0.391)	(1.029)	(0.082)	(0.176)
ISIC Code P	32.127***	3.263**	1.131	1.138
	(3.919)	(1.467)	(0.097)	(0.184)
ISIC Code Q	5.92***	5.638***	1.07	1.039
	(0.845)	(2.511)	(0.123)	(0.179)
ISIC Code R, S, T, and U	3.383***	1.516	1.024	0.994
	(0.415)	(0.796)	(0.083)	(0.184)
Health_insurance (Yes as a base)				
No health insurance	1.031	1.165	1	0.983
	(0.055)	(0.144)	(0.034)	(0.05)
Do not Know	0.82	1.017	0.96	0.934
	(0.141)	(0.428)	(0.102)	(0.177)
Pension benefit (Yes, as a base)				
No pension benefit	0.864**	1.117	1.001	1.017
	(0.045)	(0.119)	(0.038)	(0.048)
Do not Know	0.885	1.418	0.917	1.075
	(0.113)	(0.411)	(0.082)	(0.145)
Leave entitlements (Yes, as a base)				
No leave entitlements	0.917*	0.886	0.969	0.975
	(0.046)	(0.099)	(0.032)	(0.047)
Do not Know	1.010	0.877	0.92	0.867

	DUNDER_JA		DUNDER_RM	
	Secondary Educ	**Tertiary Educ**	**Secondary Educ**	**Tertiary Educ**
	(0.15)	(0.328)	(0.091)	(0.14)
Minimum wage standards (Yes, as a base)				
No minimum wage standards	0.911*	1.144	1.034	1.022
	(0.045)	(0.119)	(0.034)	(0.047)
Do not Know	0.907	0.894	1.126	0.942
	(0.106)	(0.271)	(0.084)	(0.123)
Employment Contract (Indefinite Time Work Agreement as base)				
Specific Time Employment Agreement	0.773***	0.974	1.025	1.074*
	(0.037)	(0.094)	(0.035)	(0.045)
Verbal agreements	0.611***	0.86	1.067	1.019
	(0.045)	(0.171)	(0.049)	(0.079)
No contract agrement	0.729***	1.171	1.04	1.104
	(0.047)	(0.164)	(0.045)	(0.068)
Do not Know	0.937	1.119	0.999	1.054
	(0.108)	(0.331)	(0.078)	(0.139)
Employer institution (nonprofit organization as a base)				
Profitable corporation	0.706***	1.088	1.111	1.084
	(0.057)	(0.131)	(0.074)	(0.06)
Individual business	0.342***	0.704*	1.05	1.095
	(0.03)	(0.128)	(0.072)	(0.088)
Household business	0.113***	-	1.015	0.949
	(0.028)		(0.114)	(0.325)
Others	0.622**	1.369	1.112	1.129
	(0.11)	(0.361)	(0.159)	(0.13)
Do not Know	0.742	0.978	1.159	0.626
	(0.16)	(0.654)	(0.177)	(0.207)
Spatial Characteristics				
Rural (Urban as a base)	1.039	1.492***	1.163***	1.13**
	(0.042)	(0.156)	(0.029)	(0.051)
Migration (stay as base)	1.201***	0.851*	1.039	1.065*
	(0.046)	(0.074)	(0.026)	(0.039)
Migrant (domestic worker as a base)	2.559**	-	0.958	0.745
	(1.222)		(0.343)	(0.269)
Spatial_mobility (in town as a base)	0.991	0.944	1.001	0.991
	(0.045)	(0.091)	(0.03)	(0.041)

*Notes: *significant at level 0.001, ***significant at level 0.05, ***significant at level 0.10. Standard errors are in parentheses. An odds ratio of less than 1 indicates a lower probability within the group (base), while more than 1 indicates a higher probability. If the odds ratio is converted into a coefficient, then an odds ratio of less than 1 is negative, whereas if it is more than 1, it shows a positive coefficient. Detailed explanations of ISCED and ISIC codes are in the appendix.*

Source: Estimated by authors from the raw datasets of Indonesia's Labour Force Survey 2022

Table 4 shows that from the statistical aspect of LR Chi^2, all explanatory variables in all vertical educational mismatch determinant models tested simultaneously affect the dependent variables. As for the statistical aspect of Pseudo R^2, it appears that undereducation estimated using the JA approach has a better determinant model than the RM approach. For this reason, we only discuss the estimated determinants of undereducation using the JA approach.

For workers with secondary education, the increasing age and experience make them more prone to being undereducated. Compared to workers aged 15 to 24, older workers are more likely to be undereducated. It is in line with the increase in their experience. The more work experience they have, the higher the probability of them experiencing undereducation. In addition, workers who have participated in training are also more likely to experience undereducation. In other words, undereducated Indonesian workers with secondary education rely more on aspects of experience and work training to perform their work.

Regarding educational characteristics, workers with secondary education who come from private educational institutions have a lower probability of experiencing undereducation. In the ISCED category, 6 ISCED groups have a significant probability of being undereducated when compared to ISCED code 011 (education field), namely ISCED codes 053, 073, 078, 083, 084, and 091. The existence of workers in several ISCED groups who have a greater chance of being undereducated indicates a low supply of workers in these ISCED groups. For example, ISCED code 053 is a field of physics that is relatively unpopular. As a result, employers who need workers educated in physics are 'forced' to employ undereducated workers.

Based on job and employer characteristics, most industrial sectors based on the ISIC category produce a higher probability of undereducated workers than the agriculture, forestry, and fishing sectors. This condition occurs because the years of education required for the agriculture, forestry, and fishing sectors are relatively lower than other industrial sectors. Meanwhile, workers with work contracts tend to have a lower probability of being undereducated. Then, based on the type of employer, workers in profitable institutions have a lower probability of being undereducated than workers in non-profit institutions. Based on spatial characteristics, workers who migrate to cities or other regions are more likely to be undereducated. It indicates that moving to an area with a higher level of industry does not necessarily increase the probability of workers working according to their level of education.

Compared to secondary education workers, we found heterogeneities in the determinants of undereducation in tertiary workers. In highly educated workers, females are more likely to be undereducated than men due to the more limited labour market for females. Regarding age groups, there is no significant tendency for older workers to have a higher probability of being undereducated. However, the age group of 55 to 64 years is proven to have a higher probability of being undereducated. In terms of experience, the more experience workers have, the higher the probability of being undereducated. Experience is still important for highly educated workers when carrying out their work. However, a stylised fact found in this study is that workers participating in training activities have a greater potential to be undereducated. This condition indicates a trade-off between the level of formal education and experience in the context of vertical mismatch. The determinant of undereducation that is quite dominant for highly educated workers is the type of education field. Compared to workers with educational backgrounds in education and teaching (ISCED code 011), most other educational fields are more likely to be undereducated. This condition occurs because most workers working in the education and teaching sector have a relatively higher level of education than other sectors. Therefore, workers in the scientific field of ISCED Code 011 have a lower probability of being undereducated.

Furthermore, we can not estimate the determinants of overeducation in the secondary education workforce due to the low number of observations. Table 3 shows that overeducated workers comprise only 11.22% of the total. Of this number, only a few have a secondary education background. Most overeducated workers have at least a college education (from diploma 1 to doctoral level). The odds ratios for determinants of overeducation are as follows:

Table 5: Determinants of Overeducation (Only for Workers with Tertiary Education)

	DOVER_JA Tertiary Educ	**DOVER_RM Tertiary Educ**
LR chi^2 (Prob)	5280	239
	(0.000)	(0.000)
Pseudo R^2	0.258	0.018
Obs	14,862	14,848
Constant	2.209***	0.15***
	(0.515)	(0.037)
Individual Characteristics		
Female (male as base)	0.922*	0.95
	(0.042)	(0.05)
Married (not married/divorced as base)	0.956	0.981
	(0.051)	(0.059)
Having a Child (not having a child as a base)	1.023	0.992
	(0.053)	(0.06)
Age Group (15-24 as a base)		
25 – 34 years	1.008	1.418***
	(0.075)	(0.135)
35 – 44 years	0.86*	1.488***
	(0.074)	(0.16)
45 – 54 years	0.748**	2.586***
	(0.075)	(0.304)
55 – 64 years	0.888	3.362***
	(0.115)	(0.472)
more than 64	2.712***	2.438**
	(0.779)	(0.757)
Experience Group (1-5 years as a base)		
6-10 years	0.875**	0.998
	(0.05)	(0.066)
11-15 years	0.773***	1.082
	(0.055)	(0.086)
>15 years	0.686***	0.812**
	(0.052)	(0.069)
Training (No training experience as base)	0.88**	0.977
	(0.038)	(0.048)
Education Characteristics		
Education Institution Type (Public as a base)		
Private school	0.962	1.064
	(0.042)	(0.053)
Agency School	0.242*	1.331
	(0.18)	(1.112)
Others School	-	-
ISCED 3 Digits (ISCED Code 011 as a base)		
ISCED Code 018	1.718	2.017*
	(0.655)	(0.727)
ISCED Code 021	1.864***	1.093
	(0.334)	(0.223)
ISCED Code 022	1.155	0.975
	(0.136)	(0.126)
ISCED Code 023	1.557**	1.392**
	(0.237)	(0.229)
ISCED Code 028	1.967***	0.988
	(0.376)	(0.228)
ISCED Code 030	3.5***	1.129

	DOVER_JA Tertiary Educ	DOVER_RM Tertiary Educ
	(0.6)	(0.208)
ISCED Code 031	2.772***	0.974
	(0.248)	(0.101)
ISCED Code 032	0.55	-
	(0.543)	
ISCED Code 038	1.775*	1.069
	(0.565)	(0.419)
ISCED Code 041	2.965***	1.081
	(0.235)	(0.098)
ISCED Code 042	2.878***	1.169
	(0.335)	(0.151)
ISCED Code 048	1.977	-
	(2.017)	
ISCED Code 050	-	-
ISCED Code 051	1.699**	0.809
	(0.289)	(0.181)
ISCED Code 052	1.509	1.413
	(0.403)	(0.417)
ISCED Code 053	1.12	0.996
	(0.224)	(0.249)
ISCED Code 054	1.154	1.251
	(0.211)	(0.245)
ISCED Code 061	2.325***	1.02
	(0.213)	(0.11)
ISCED Code 068	0.897	0.796
	(0.688)	(0.863)
ISCED Code 070	0.199	1.221
	(0.227)	(1.376)
ISCED Code 071	1.221*	1.042
	(0.137)	(0.143)
ISCED Code 072	0.944	0.85
	(0.145)	(0.175)
ISCED Code 073	1.115	1.241
	(0.133)	(0.18)
ISCED Code 078	0.47**	0.989
	(0.13)	(0.336)
ISCED Code 079	1.668**	0.9
	(0.271)	(0.188)
ISCED Code 080	-	4.155
		(5.913)
ISCED Code 081	3.021***	0.908
	(0.39)	(0.134)
ISCED Code 082	3.005**	1.194
	(1.286)	(0.518)
ISCED Code 083	1.869**	0.946
	(0.579)	(0.35)
ISCED Code 084	2.676	-
	(2.992)	
ISCED Code 088	-	-
ISCED Code 090	0.737	1.32
	(0.364)	(0.75)
ISCED Code 091	0.946	0.92
	(0.102)	(0.123)

	DOVER_JA Tertiary Educ	DOVER_RM Tertiary Educ
ISCED Code 098	6.706	1.517
	(7.887)	(1.778)
ISCED Code 101	2.039***	0.671
	(0.442)	(0.196)
ISCED Code 102	5.131*	0.787
	(4.913)	(0.875)
ISCED Code 103	0.473	6.505
	(0.695)	(9.466)
ISCED Code 104	0.921	0.495
	(0.25)	(0.22)
Job and Employer Characteristics		
Industry Sectors (ISIC Code A as a base)		
ISIC Code B	0.885	1.017
	(0.207)	(0.237)
ISIC Code C	0.876	0.739
	(0.179)	(0.149)
ISIC Code D	0.534**	0.748
	(0.138)	(0.216)
ISIC Code E	1.33	0.55
	(0.584)	(0.263)
ISIC Code F	0.584**	0.547**
	(0.127)	(0.125)
ISIC Code G	1.106	0.73
	(0.225)	(0.145)
ISIC Code H	1.155	0.747
	(0.263)	(0.167)
ISIC Code I	0.839	0.774
	(0.194)	(0.18)
ISIC Code J	0.221***	0.757
	(0.048)	(0.17)
ISIC Code K	0.647**	0.785
	(0.13)	(0.155)
ISIC Code L	0.403***	0.733
	(0.115)	(0.236)
ISIC Code M and N	0.31***	0.759
	(0.065)	(0.162)
ISIC Code O	1.009	0.887
	(0.239)	(0.208)
ISIC Code P	0.11***	0.684*
	(0.023)	(0.138)
ISIC Code Q	0.166***	0.76
	(0.036)	(0.167)
ISIC Code R, S, T, and U	0.493**	0.658*
	(0.112)	(0.156)
Health_insurance (Yes as a base)		
No health insurance	0.733***	1.169**
	(0.045)	(0.081)
Do not Know	0.967	1.539*
	(0.218)	(0.365)
Pension benefit (Yes, as a base)		
No pension benefit	1.114*	1.004
	(0.062)	(0.066)
Do not Know	1.059	0.971

	DOVER_JA Tertiary Educ	DOVER_RM Tertiary Educ
	(0.171)	(0.181)
Leave entitlements (Yes, as a base)		
No leave entitlement	1.133**	0.986
	(0.066)	(0.066)
Do not Know	1.039	1.241
	(0.2)	(0.265)
Minimum wage standards (Yes, as a base)		
No minimum wage standards	0.932	0.961
	(0.052)	(0.062)
Do not Know	1.025	0.912
	(0.162)	(0.163)
Employment Contract (Indefinite Time Work Agreement as base)		
Specific Time Employment Agreement	1.053	0.988
	(0.053)	(0.057)
Verbal agreements	1.262**	0.954
	(0.121)	(0.102)
No contract agreement	1.048	1.028
	(0.078)	(0.086)
Do not Know	1.187	1.047
	(0.185)	(0.181)
Employer institution (nonprofit organization as a base)		
Profitable corporation	1.271***	0.928
	(0.085)	(0.071)
Individual business	2.473***	0.867
	(0.24)	(0.096)
Household business	4.042***	0.703
	(1.77)	(0.347)
Others	0.901	0.918
	(0.138)	(0.145)
Do not Know	2.304**	1.618
	(0.829)	(0.599)
Spatial Characteristics		
Rural (Urban as a base)	0.845**	1.36***
	(0.047)	(0.082)
Migration (stay as base)	0.989	1.001
	(0.044)	(0.051)
Migrant (domestic worker as a base)	1.532	1.395
	(0.681)	(0.604)
Spatial_mobility (in town as a base)	1.016	1.034
	(0.05)	(0.06)

*Notes: *significant at level 0.001, ***significant at level 0.05, ***significant at level 0.10. Standard errors are in parentheses.*

Source: Estimated by authors from the raw datasets of Indonesia's Labour Force Survey 2022

The determinant overeducation model with the JA approach has a better GOF than the RM approach (see Pseudo R^2 value). Thus, based on the JA approach, we will discuss the determinants of overeducation and undereducation. In this context, Hartog (2000) agreed that the JA normative approach would be relatively more accurate in estimating vertical educational mismatch. If estimated using the RM approach, the number of overeducated workers may be biased due to differences between secondary and tertiary education workers.

If analysed only based on a highly educated workforce (tertiary education), this study found that female workers have a lower probability of being undereducated and are more susceptible to experiencing overeducation. This finding is relatively in line with several previous studies, including Vahey (2000), Senkrua (2015), and McGuinness et al. (2018). These findings show that higher education is needed for females to enter the world of work. The reason is that female employment opportunities in the industrial sector tend to be more limited. It confirms the relevance of Frank's (1978) differential overqualification theory. This theory predicts that females will have a greater probability of experiencing educational mismatch. The theory also predicts that if a married woman follows her husband to an area with a lower labour market, her probability of experiencing overeducation or undereducation will be higher (Büchel & Battu, 2003). However, in this study, marital status and having a child were not proven to be determinants of overeducation or undereducation.

This study found that age is one of the determinants of vertical educational mismatch. The probability of being undereducated becomes higher as the workers age. One of the reasons for the increasing probability of workers experiencing vertical educational mismatch as age increases is low participation in higher education. This study is in line with several studies, including Devillanova (2013) and Leuven & Oosterbeek (2011). Their study also found that age affects overeducation. As age increases, experience and competence can also increase to determine the probability of being overeducated. Additionally, according to Ordine & Rose (2017), increasing access to higher education would reduce vertical educational mismatch.

On the other hand, experience can increase a worker's probability of experiencing undereducation and reduce the probability of experiencing overeducation. These findings confirm the stepping stone hypothesis based on human capital and career mobility theory. This hypothesis states that overeducation decreases as work experience increases and career mobility increases. Several studies also found that experience can reduce overeducation, namely Kupets (2015) and Senkrua (2015). This study is similar to Chaya et al. (2013), who stated that career mobility (characterised by experience) can reduce overeducation.

Workers who have received training have a lower probability of experiencing undereducation. On the contrary, such training can increase a worker's probability of experiencing overeducation. This condition can be caused by the shifting phenomenon, where an individual prefers a job that does not suit their field and level of education. After getting the job, the workers will need additional training because the skills they get from education differ from the job they are currently doing. However, this needs to be studied more by examining the determinants of horizontal educational mismatch. This study does not focus in that direction.

The determinant of undereducation based on the two digits of the ISCED code shows that workers in all scientific fields are more likely to be undereducated than ISCED Code 01 (teacher education). This condition is caused by the high composition of undereducated workers in the Indonesian wage sector. These findings are relatively similar to previous studies by Carroll & Tani (2015) and Zheng et al. (2021). Their research also found variations in the probability of overeducation based on the field of study. The level of overeducated workers in a field of study shows its relevance to the industrial world.

Furthermore, this study confirms the relevance of CMT and matching theory that job and employer characteristics can determine vertical educational mismatch. Workers in companies that provide health insurance, retirement benefits, and standard salary policies are more likely to be overeducated. Companies that implement these employment policies tend to be larger. Therefore, career mobility is also transparent, so highly educated workers are still willing to work at the company even if they must be overeducated. Several previous studies, including Karakaya et al. (2007), Belfield (2010), Zakariya & Noor (2014), and Ege & Erdil (2023), have also found this fact.

Finally, this study found that spatial characteristics affect vertical educational mismatch. Workers in rural areas have a higher probability of being undereducated and a lower probability of being overeducated. The lower average education of workers in rural areas causes this. Besides, the high probability of workers being undereducated in rural areas is also caused by the lack of highly educated workers. Workers with higher education in Indonesia tend to migrate and have spatial mobility to find work that matches their educational level.

4.3. The wage effect of vertical educational mismatch in Indonesia's waged sector

Before examining the wage effect of vertical educational mismatch, we first describe the general conditions of the variables that will be tested in this study as follows:

Table 6: Descriptive Statistics of Variables in Wage Equation

	Mean	**Std.Dev**	**Min**	**Max**	**Obs**	**Population**
Male in Urban:						
YUNDER_JA	1.403	2.156	0	10	32,559	14,620,579
YOVER_JA	0.392	1.121	0	7	32,559	14,620,579
YREQ_JA	12.704	1.429	12	16	32,559	14,620,579
DUNDER_JA	0.341	0.474	0	1	32,559	14,620,579
DOVER_JA	0.126	0.332	0	1	32,559	14,620,579
EDUC	11.694	2.894	6	22	32,559	14,620,579
EXPER	8.757	8.958	0	62	32,559	14,620,579
AGE	36.959	11.52	15	89	32,559	14,620,579
Female in Urban:						
YUNDER_JA	1.094	1.985	0	10	18,240	7,897,877
YOVER_JA	0.571	1.330	0	6	18,240	7,897,877
YREQ_JA	13.005	1.688	12	16	18,240	7,897,877
DUNDER_JA	0.277	0.447	0	1	18,240	7,897,877
DOVER_JA	0.171	0.376	0	1	18,240	7,897,877
EDUC	12.482	2.992	6	22	18,240	7,897,877
EXPER	6.594	7.785	0	58	18,240	7,897,877
AGE	33.285	11.229	15	86	18,240	7,897,877
Male in Rural :						
YUNDER_JA	2.349	2.530	0	10	18,174	4,486,352
YOVER_JA	0.154	0.730	0	6	18,174	4,486,352
YREQ_JA	12.469	1.216	12	16	18,174	4,486,352
DUNDER_JA	0.522	0.500	0	1	18,174	4,486,352
DOVER_JA	0.048	0.213	0	1	18,174	4,486,352
EDUC	10.275	2.932	6	22	18,174	4,486,352
EXPER	7.340	7.885	0	61	18,174	4,486,352
AGE	34.877	11.028	15	98	18,174	4,486,352
Female in Rural:						
YUNDER_JA	1.952	2.433	0	10	7,774	2,091,289

	Mean	**Std.Dev**	**Min**	**Max**	**Obs**	**Population**
YOVER_JA	0.234	0.895	0	4	7,774	2,091,289
YREQ_JA	13.007	1.711	12	16	7,774	2,091,289
DUNDER_JA	0.446	0.497	0	1	7,774	2,091,289
DOVER_JA	0.068	0.252	0	1	7,774	2,091,289
EDUC	11.288	3.188	6	18	7,774	2,091,289
EXPER	5.295	6.842	0	56	7,774	2,091,289
AGE	31.76	10.859	15	80	7,774	2,091,289

Notes: The units of YUNDER, YOVER, YREQ, EDUC, EXPER, and AGE are years, while DUNDER and DOVER are categorical variables with values 0 and 1

Source: Estimated by authors from the raw datasets of Indonesia's Labour Force Survey 2022

Table 6 shows that the average education level of female workers is higher than that of male workers in urban and rural areas. The workforce composition in Indonesia's wage sector is 66.1% male and 33.9% female. As a result, female job qualifications require a higher level of education than men. It has been proven that the required years of education (YREQ) for females are higher than for males in urban and rural areas. Meanwhile, the average age of male and female workers in urban and rural areas is relatively close, around 31 to 36 years. This condition shows the ideal working age and should be more productive.

This study correlates all explanatory variables to ensure that no multicollinearity appears:

Table 7: Correlation Matrix

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
YUNDER_JA (1)	1							
YOVER_JA (2)	-0.234	1						
YREQ_JA (3)	0.032	-0.080	1					
DUNDER_JA (4)	0.899	-0.260	0.099	1				
DOVER_JA (5)	-0.250	0.937	0.008	-0.278	1			
EDUC (6)	-0.817	0.492	0.436	-0.717	0.524	1		
EXPER (7)	0.135	-0.024	0.100	0.126	-0.018	-0.061	1	
AGE (8)	0.237	-0.010	0.085	0.209	-0.003	-0.140	0.588	1

Source: Estimated by authors from the raw datasets of Indonesia's Labour Force Survey 2022

Table 7 shows a high correlation between variables that occurs between DUNDER_JA and YUNDER_JA. YOVER_JA and DOVER_JA also have a very high correlation, meaning that the vertical educational mismatch, defined as surplus and deficit years of education, and the dummy variables of overeducation and undereducation tend to be identical. Meanwhile, EDUC with YUNDER_JA experienced a very high negative correlation. However, this does not indicate a multicollinearity problem because, from the correlation matrix above, it will be arranged into two models: ORU and VV. Overall, we confirm no multicollinearity problems in the models. We also confirm that our data for all regression models is normally distributed. We employ the histogram and scatter plot to check the normal distribution and heteroscedasticity test, but we have not shown the results here for brevity. This study's model was also homoscedastic. However, due to the relatively large number of observations, there is doubt that the error term distribution is constant. For this reason, we use cluster-robust standard errors in each regression model.

The results of the ORU and VV model estimates categorised by gender and spatial characteristics (urban and rural) are as follows:

Table 8: ORU and VV Wage Equation Based on Gender and Spatial (JA Approach)

	I	II	III	IV
Male:				
Constant	7.301***	8.555***	7.771***	8.613***
	(0.121)	(0.145)	(0.088)	(0.086)
YOVER_JA	0.114***	0.061***	-	-
	(0.008)	(0.007)		
YUNDER_JA	-0.066***	-0.032***	-	-
	(0.007)	(0.004)		
YREQ_JA	0.163***	0.048***	-	-
	(0.013)	(0.013)		
DOVER_JA	-	-	0.006	0.103***
			(0.027)	(0.03)
DUNDER_JA	-	-	0.247***	0.068**
			(0.021)	(0.025)
EDUC	-	-	0.124***	0.043***
			(0.01)	(0.007)
EXPER	0.011***	0.004**	0.011***	0.004**
	(0.002)	(0.002)	(0.002)	(0.002)
AGE	0.007***	0.009***	0.008***	0.009***
	(0.001)	(0.001)	(0.001)	(0.001)
F (Prob)	202	106.4	209.86	118.83
	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.223	0.054	0.209	0.055
Root MSE	0.621	0.611	0.626	0.611
Obs	32,559	18,174	32,559	18,174
Female:				
Constant	8.005***	8.69***	8.058***	8.537***
	(0.135)	(0.203)	(0.101)	(0.149)
YOVER_JA	0.145***	0.098***	-	-
	(0.007)	(0.014)		
YUNDER_JA	-0.087***	-0.047***	-	-
	(0.006)	(0.009)		
YREQ_JA	0.091***	0.016	-	-
	(0.013)	(0.014)		
DOVER_JA	-	-	0.245***	0.309***
			(0.038)	(0.043)
DUNDER_JA	-	-	0.017	-0.04
			(0.032)	(0.04)
EDUC	-	-	0.086***	0.029**
			(0.01)	(0.01)
EXPER	0.022***	0.017***	0.022***	0.017***
	(0.004)	(0.003)	(0.004)	(0.003)

	I	II	III	IV
AGE	0.002*	0.002	0.002	0.001
	(0.002)	(0.002)	(0.002)	(0.003)
F (Prob)	108.81	50.4	111.62	48.01
	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.205	0.060	0.208	0.060
Root MSE	0.705	0.725	0.703	0.725
Obs	18,240	7,774	18,240	7,774

*Notes: *significant at level 0.001, ***significant at level 0.05, ***significant at level 0.10. Cluster robust standard error between provinces are in parentheses. Dependent variable = LnWage per hour. Column I is the estimation result of the ORU model in urban areas. Column II results from the ORU model estimation for the rural area. Column III results from the VV model estimation for the urban area. Column IV is the estimation result of the VV model for the rural area*

Source: Estimated by authors from the raw datasets of Indonesia's Labour Force Survey 2022

The ORU model estimation results show a positive return to all overeducated workers in both urban and rural areas. This positive return to overeducation shows that even though they work below their educational level, the additional education of overeducated workers still brings economic value to their wages. However, the return for overeducated male workers in urban areas is lower than that of required-match workers. The difference between these returns is only 0.019 (0.63 - 0.114) or 1.9%, implying that overeducated male workers in urban area companies get a lower return on education, 1.9%, compared to their coworkers. As an illustration, two professional accountants have bachelor's and master's degrees, respectively. These two people will earn relatively similar wages. However, because a professional accountant with a master's degree has made a more significant educational investment than an accountant with a bachelor's degree, the rate of return on their education is lower. Meanwhile, because the return is positive, the increase in the number of years of education of overeducated workers still positively impacts wages.

For female workers in urban areas, overeducation will produce higher returns than the required match. Likewise, overeducated female workers in rural areas still get positive returns whose value is higher than the required match. This condition indicates that higher education is needed for females to get jobs with higher wage levels, which is confirmed by the higher incidence of overeducation among females than males. Overeducated females comprise 14.2% of the population, while overeducated men comprise 9.78%. Moreover, the logit estimation results also proved that females have a higher probability of experiencing overeducation (see Table 5).

On the other hand, the VV model shows that overeducated male workers in urban areas earn no penalty or wage premium. Meanwhile, there is a wage premium for undereducated male workers in urban and rural areas. In urban areas, undereducated male workers earn 24.7% higher wages than match workers at the same education level. In rural areas, undereducated male workers earn 6.8% more wages. Meanwhile, overeducated female workers in urban areas earn 24.5% more wages than other workers at the same education level. In rural areas, the wages earned by overeducated female workers are 30.9% greater than those at the same educational level.

Furthermore, to examine the impact of vertical educational mismatch on wages more comprehensively, the following are the estimation results of the ORU model categorised based on the first digit ISCED code:

Table 9: ORU Model Estimation Results Based on First Digit ISCED (JA Approach)

	ISCED Code 01	**ISCED Code 02**	**ISCED Code 03**	**ISCED Code 04**	**ISCED Code 05**	**ISCED Code 06**
Urban:						
Constant	5.706***	8.478***	6.861***	6.506***	7.676***	6.554***
	(0.558)	(0.206)	(0.177)	(0.156)	(0.31)	(0.306)
YOVER_JA	0.257***	0.113***	0.122***	0.132***	0.138***	0.118***
	(0.029)	(0.01)	(0.011)	(0.008)	(0.014)	(0.011)
YUNDER_JA	0.495	-0.07**	-0.138***	-0.159***	-0.079**	-0.133***
	(0.322)	(0.024)	(0.022)	(0.02)	(0.026)	(0.027)
YREQ_JA	0.189***	0.029	0.187***	0.222***	0.112***	0.198***
	(0.036)	(0.018)	(0.019)	(0.017)	(0.026)	(0.027)
EXPER	0.018***	0.012***	0.016***	0.022***	0.018***	0.025***
	(0.004)	(0.003)	(0.002)	(0.002)	(0.002)	(0.004)
AGE	0.012***	0.013***	0.008***	0.004	0.011***	0.013***
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)
F (Prob)	66.26	117.09	111.48	214.32	77.07	91.86
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R2	0.127	0.13	0.176	0.313	0.161	0.336
Root MSE	0.801	0.761	0.635	0.656	0.662	0.65
Obs	1796	1720	10545	6101	5264	2313
Rural:						
Constant	5.629**	9.13***	9.063***	8.046***	8.895***	8.368***
	(1.626)	(0.369)	(0.307)	(0.264)	(0.253)	(0.327)
YOVER_JA	0.283**	0.094***	0.058***	0.097***	0.081**	0.052**
	(0.097)	(0.021)	(0.015)	(0.014)	(0.027)	(0.018)
YUNDER_JA	-0.026	-0.028	-0.022	-0.097**	-0.03	0.027
	(0.177)	(0.025)	(0.028)	(0.036)	(0.023)	(0.039)
YREQ_JA	0.157	-0.00145	-0.016	0.077***	-0.015	0.017
	(0.102)		(0.026)	(0.021)	(0.02)	(0.031)
EXPER	0.008	0.01	0.009***	0.029***	0.011**	0.017
	(0.009)	(0.007)	(0.003)	(0.004)	(0.004)	(0.011)
AGE	0.024***	0.02***	0.014***	0.004	0.018***	0.023***
	(0.007)	(0.004)	(0.002)	(0.004)	(0.003)	(0.006)
F (Prob)	40.17	21.26	86.93	52.93	21.49	41.31
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R2	0.121	0.122	0.069	0.154	0.106	0.108
Root MSE	0.83	0.771	0.655	0.649	0.691	0.631
Obs	1021	658	5508	1223	2450	670

*Notes: *significant at level 0.001, ***significant at level 0.05, ***significant at level 0.10. Cluster robust standard errors between provinces are in parentheses. Dependent variable= Ln Wage per hour. ISCED Code 01 = Education, ISCED Code 02 = Arts and Humanities, and so on (full ISCED Code descriptions are in Appendix)*

Source: Estimated by authors from the raw datasets of Indonesia's Labour Force Survey 2022

Table 9: ORU Model Estimation Results Based on First Digit ISCED (JA Approach) *(cont.)*

	ISCED Code 07	**ISCED Code 08**	**ISCED Code 09**	**ISCED Code 10**	**Unidentified ISCED**	**Total**
Urban:						
Constant	6.593***	7.06***	6.77***	6.335***	7.681***	7.697***
	(0.185)	(0.34)	(0.207)	(0.495)	(0.165)	(0.131)
YOVER_JA	0.136***	0.099***	0.143***	0.141***	0.149***	0.122***
	(0.01)	(0.019)	(0.019)	(0.022)	(0.016)	(0.008)
YUNDER_JA	-0.136***	-0.12**	-0.121***	-0.163**	-0.051***	-0.07***
	(0.02)	(0.052)	(0.025)	(0.05)	(0.006)	(0.007)
YREQ_JA	0.221***	0.166***	0.153***	0.221***	0.133***	0.121***
	(0.017)	(0.028)	(0.015)	(0.047)	(0.014)	(0.014)
EXPER	0.017***	0.019**	0.012**	0.013	0.011***	0.015***
	(0.003)	(0.007)	(0.005)	(0.007)	(0.002)	(0.002)
AGE	0.008***	0.009**	0.02***	0.012**	0.002**	0.007***
	(0.002)	(0.003)	(0.004)	(0.006)	(0.001)	(0.001)
F (Prob)	428.81	29.26	67.73	127.21	74.28	129.62
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R2	0.284	0.25	0.292	0.241	0.099	0.1908
Root MSE	0.593	0.689	0.619	0.611	0.643	0.6738
Obs	7210	805	1226	942	12877	50799
Rural:						
Constant	7.322***	8.325***	8.156***	7.435***	8.517***	8.94***
	(0.223)	(0.446)	(0.384)	(0.884)	(0.157)	(0.138)
YOVER_JA	0.099***	0.057**	0.069**	0.069	0.023	0.066***
	(0.028)	(0.019)	(0.023)	(0.046)	(0.031)	(0.009)
YUNDER_JA	-0.081***	0.004	-0.019	-0.027	-0.039***	-0.034***
	(0.016)	(0.037)	(0.052)	(0.088)	(0.007)	(0.006)
YREQ_JA	0.141***	0.036	0.051*	0.102	0.059***	0.006
	(0.018)	(0.032)	(0.027)	(0.086)	(0.014)	(0.012)
EXPER	0.013***	0.02**	0.024**	0.003	0.008***	0.011***
	(0.003)	(0.01)	(0.011)	(0.014)	(0.002)	(0.002)
AGE	0.011***	0.02***	0.012	0.023**	0.004**	0.009***
	(0.003)	(0.006)	(0.009)	(0.01)	(0.002)	(0.002)
F (Prob)	43.51	27.27	13.94	7.97	20.74	90.07
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R2	0.144	0.166	0.141	0.162	0.026	0.05
Root MSE	0.571	0.662	0.574	0.586	0.675	0.682
Obs	2076	498	312	227	11305	25948

*Notes: *significant at level 0.001, ***significant at level 0.05, ***significant at level 0.10. Cluster robust standard errors between provinces are in parentheses. Dependent variable= Ln Wage per hour. ISCED Code 01 = Education, ISCED Code 02 = Arts and Humanities, and so on (full ISCED Code descriptions are in Appendix)*

Source: Estimated by authors from the raw datasets of Indonesia's Labour Force Survey 2022

Table 9 shows that 12,877 samples in urban areas and 11,305 samples in rural areas fall into the ISCED unidentified criteria. ISCED cannot identify the field of study of these samples because we included samples with an education level below high school who did not yet have a major in education. Even though in ISCED, there is code 00, which is a generic program and qualifications, we did not use this categorisation. The reason is that our study only focuses on vertical educational mismatch. It does not discuss horizontal educational mismatch.

In urban areas, overeducation in all fields of study groups has positive returns. At the same time, undereducation has negative returns (except ISCED Code 01).

Overeducated workers still get economic benefits from the surplus schooling they have. In urban areas, an additional year of education for overeducated workers will earn a return of 12.2% on wages. It is slightly higher than the return to required education, which is 12.1%. It shows employers still appreciate highly educated workers, even if overeducated. However, there are seven fields of study where the rate of return to required education is higher than overeducation. Overeducated workers with fields of study in social sciences, journalism, and information (ISCED Code 03) experience a wage penalty of 6.5% (0.187 – 0.122 = 0.065). Overeducated workers majoring in business, administration, and law (ISCED Code 04) have a wage penalty of 9%, majors in information and communication (ICT) (ISCED Code 06) of 8%, majors of engineering, manufacturing, and construction (ISCED Code 07) of 8.5%, majoring in agriculture, forestry, fisheries, and veterinary (ISCED Code 08) at 6.7%, health and welfare (ISCED 09) at 1%, and services (ISCED 10) at 8%.

In rural areas, there are positive returns to overeducation in most fields of study. In general, overeducated workers in rural areas earn a return of 6.6% for every additional year of their education level. However, the positive return to overeducated workers in the ISCED code 10 and unidentified ISCED categories is insignificant. The return to required education, which appears to be quite large in rural areas, is in the ISED code 07 category. For this category, the ORU model captures the wage penalty for overeducated workers at 4.2% lower than their coworkers with matched status. Apart from the ISCED code 07 category, the ORU model cannot capture the wage penalty for overeducated workers in rural areas because the return to required education in most ISCED categories is not greater than the return to overeducation. It shows that in rural areas, a higher level of education is required to get a job with a high wage. It is inseparable from the lack of demand for highly educated workers in rural areas.

Furthermore, the results of the VV model estimation based on ISCED are as table 10. In general, overeducated workers in urban areas earn 12.1%, while undereducated workers earn 13.3% higher wages than matched workers at the same educational level. The contribution of education to wages is still relatively low, only 9.9%, indicating the low external education efficiency. However, there is variation in the educational vertical mismatch coefficient on wages between fields of study as measured by the JA approach. The results shown by the VV model are relatively similar to those of the ORU model. In urban areas, there is a wage penalty for overeducated workers in seven fields of study: social sciences, journalism, and information (ISCED Code 03), business sciences, administration and law (ISCED Code 04), information sciences and communication technologies (ICTs) (ISCED Code 06), engineering, manufacturing and construction sciences (ISCED 07), and agricultural, forestry, fisheries, and veterinary sciences (ISCED 08).

Nevertheless, the wage penalty for overeducated workers in the ISCED Code 08 and 09 fields of study in the VV model is insignificant. Overeducated ISCED 03 workers in urban areas earn 20.1% lower wages than match workers at the same education level. Meanwhile, undereducated workers under ISCED Code 03 receive a wage premium 20.2% higher than match workers at the same education level.

On the other hand, in general, overeducated workers in rural areas will get a wage premium, while education level only has a positive effect on wages of 2.1%. In rural areas, the contribution of education in increasing wages tends to be lower than in urban areas. The reason is that most Indonesian workers in rural areas still have low education. The average years of education for male workers in rural areas are only 10.2 years (equivalent to grade 1 upper secondary students). In comparison, female workers are 11.2 years old (equivalent to grade 2 upper secondary students). A low level of education will result in low returns, as predicted by the human capital theory from Becker (1992), which states that the higher the level of education, the higher the individual's wage.

Table 10: OLS Wage Equation VV Model Estimation Results (JA Approach)

	ISCED Code 01	**ISCED Code 02**	**ISCED Code 03**	**ISCED Code 04**	**ISCED Code 05**	**ISCED Code 06**
Urban:						
Constant	5.706***	8.478***	6.861***	6.506***	7.676***	6.554***
	(0.558)	(0.206)	(0.177)	(0.156)	(0.31)	(0.306)
YOVER_JA	0.257***	0.113***	0.122***	0.132***	0.138***	0.118***
	(0.029)	(0.01)	(0.011)	(0.008)	(0.014)	(0.011)
YUNDER_JA	0.495	-0.07**	-0.138***	-0.159***	-0.079**	-0.133***
	(0.322)	(0.024)	(0.022)	(0.02)	(0.026)	(0.027)
YREQ_JA	0.189***	0.029	0.187***	0.222***	0.112***	0.198***
	(0.036)	(0.018)	(0.019)	(0.017)	(0.026)	(0.027)
EXPER	0.018***	0.012***	0.016***	0.022***	0.018***	0.025***
	(0.004)	(0.003)	(0.002)	(0.002)	(0.002)	(0.004)
AGE	0.012***	0.013***	0.008***	0.004	0.011***	0.013***
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)
F (Prob)	66.26	117.09	111.48	214.32	77.07	91.86
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R2	0.127	0.13	0.176	0.313	0.161	0.336
Root MSE	0.801	0.761	0.635	0.656	0.662	0.65
Obs	1796	1720	10545	6101	5264	2313
Rural:						
Constant	5.629**	9.13***	9.063***	8.046***	8.895***	8.368***
	(1.626)	(0.369)	(0.307)	(0.264)	(0.253)	(0.327)
YOVER_JA	0.283**	0.094***	0.058***	0.097***	0.081**	0.052**
	(0.097)	(0.021)	(0.015)	(0.014)	(0.027)	(0.018)
YUNDER_JA	-0.026	-0.028	-0.022	-0.097**	-0.03	0.027
	(0.177)	(0.025)	(0.028)	(0.036)	(0.023)	(0.039)
YREQ_JA	0.157		-0.016	0.077***	-0.015	0.017
	(0.102)	-0.00145	(0.026)	(0.021)	(0.02)	(0.031)
EXPER	0.008	0.01	0.009***	0.029***	0.011**	0.017
	(0.009)	(0.007)	(0.003)	(0.004)	(0.004)	(0.011)
AGE	0.024***	0.02***	0.014***	0.004	0.018***	0.023***
	(0.007)	(0.004)	(0.002)	(0.004)	(0.003)	(0.006)
F (Prob)	40.17	21.26	86.93	52.93	21.49	41.31
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R2	0.121	0.122	0.069	0.154	0.106	0.108
Root MSE	0.83	0.771	0.655	0.649	0.691	0.631
Obs	1021	658	5508	1223	2450	670

*Notes:***significant at level 0.001,* ****significant at level 0.05,* ****significant at level 0.10. The cluster robust standard error between provinces is in parentheses. Dependent variable = Ln Wage per hour*

Source: Estimated by authors from the raw datasets of Indonesia's Labour Force Survey 2022

Table 10: OLS Wage Equation VV Model Estimation Results (JA Approach) *(cont.)*

	ISCED Code 07	**ISCED Code 08**	**ISCED Code 09**	**ISCED Code 10**	**Unidentified ISCED**	**Total**
Urban:						
Constant	6.593***	7.06***	6.77***	6.335***	7.681***	7.697***
	(0.185)	(0.34)	(0.207)	(0.495)	(0.165)	(0.131)
YOVER_JA	0.136***	0.099***	0.143***	0.141***	0.149***	0.122***
	(0.01)	(0.019)	(0.019)	(0.022)	(0.016)	(0.008)
YUNDER_JA	-0.136***	-0.12**	-0.121***	-0.163**	-0.051***	-0.07***
	(0.02)	(0.052)	(0.025)	(0.05)	(0.006)	(0.007)
YREQ_JA	0.221***	0.166***	0.153***	0.221***	0.133***	0.121***
	(0.017)	(0.028)	(0.015)	(0.047)	(0.014)	(0.014)
EXPER	0.017***	0.019**	0.012**	0.013	0.011***	0.015***
	(0.003)	(0.007)	(0.005)	(0.007)	(0.002)	(0.002)
AGE	0.008***	0.009**	0.02***	0.012**	0.002**	0.007***
	(0.002)	(0.003)	(0.004)	(0.006)	(0.001)	(0.001)
F (Prob)	428.81	29.26	67.73	127.21	74.28	129.62
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R2	0.284	0.25	0.292	0.241	0.099	0.1908
Root MSE	0.593	0.689	0.619	0.611	0.643	0.6738
Obs	7210	805	1226	942	12877	50799
Rural:						
Constant	7.322***	8.325***	8.156***	7.435***	8.517***	8.94***
	(0.223)	(0.446)	(0.384)	(0.884)	(0.157)	(0.138)
YOVER_JA	0.099***	0.057**	0.069**	0.069	0.023	0.066***
	(0.028)	(0.019)	(0.023)	(0.046)	(0.031)	(0.009)
YUNDER_JA	-0.081***	0.004	-0.019	-0.027	-0.039***	-0.034***
	(0.016)	(0.037)	(0.052)	(0.088)	(0.007)	(0.006)
YREQ_JA	0.141***	0.036	0.051*	0.102	0.059***	0.006
	(0.018)	(0.032)	(0.027)	(0.086)	(0.014)	(0.012)
EXPER	0.013***	0.02**	0.024**	0.003	0.008***	0.011***
	(0.003)	(0.01)	(0.011)	(0.014)	(0.002)	(0.002)
AGE	0.011***	0.02***	0.012	0.023**	0.004**	0.009***
	(0.003)	(0.006)	(0.009)	(0.01)	(0.002)	(0.002)
F (Prob)	43.51	27.27	13.94	7.97	20.74	90.07
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R2	0.144	0.166	0.141	0.162	0.026	0.05
Root MSE	0.571	0.662	0.574	0.586	0.675	0.682
Obs	2076	498	312	227	11305	25948

*Notes:*significant at level 0.001, ***significant at level 0.05, ***significant at level 0.10. The cluster robust standard error between provinces is in parentheses. Dependent variable = Ln Wage per hour*

Source: Estimated by authors from the raw datasets of Indonesia's Labour Force Survey 2022

After estimating the ORU and VV models, this study found that the effect of educational vertical mismatch, whether gender, spatial conditions, and fields of study, determines overeducation or undereducation on wages. Overeducated workers, both male and female, get positive returns in urban and rural areas. Overeducated workers get premium wages (except for overeducated male workers in urban areas), as do undereducated workers (VV model estimation results). This finding indicates that highly educated workers will get higher wages than matched workers at the same level of education. However, on the other hand, workers with low education will also get higher wages if they are given many assignments. However, this wage premium is a loss borne by employers (Duncan & Hoffman, 1981; Hartog & Oosterbeek, 1988).

These findings simultaneously confirm the relevance of human capital theory (Becker, 1992) and work assignment theory (Sattinger, 1993). According to human capital theory, the greater a person's investment in human capital (in this case, characterised by years of education), the higher the wage they will get. Therefore, overeducated workers get higher wages. On the other hand, work assignment theory indicates that work assignments determine wages, so undereducated workers will also receive higher wages than matched workers at the same educational level. The existence of positive returns on overeducation and negative returns on undereducation in the ORU model in this study is relatively similar to several previous studies, including Zakariya (2014) in Malaysia, Grunau & Pecoraro (2017) in Germany, Clark et al. (2017) in the US, and others.

This study is relatively contradictory to many previous studies, including Diem & Wolter (2014), Iriondo & Amaral (2016), Park & Jang (2017), Johnes (2019), Schweri et al. (2020), Carmichael et al. (2021), and Sun & Kim (2022). Using the VV model, their study found a wage penalty for overeducated workers. Meanwhile, this study found a wage premium for overeducated workers with the VV model. Spatial conditions and fields of study can cause this difference because this study also found wage penalties for overeducated workers in several fields of study groups in urban areas (see Table 9). However, in rural areas, this study found no wage penalties for overeducated workers in all ISCED categories.

5. Conclusion

This study found that as many as 49.2% of workers in the Indonesian wage sector work at jobs that do not match their educational level. As many as 37.9% of the workers are undereducated, while the rest, 11.2%, are overeducated. Vertical educational mismatch causes are individual characteristics, educational and fields of study, job and employer characteristics, and spatial characteristics. In Indonesia's wage sector, this vertical educational mismatch impacts wages. Using the ORU and VV models, this study found that the wage effect of vertical educational mismatch depends on gender, fields of study, and spatial conditions. In general, overeducated workers in urban areas earn higher returns than those in rural areas. Based on the ORU model, overeducated workers experience a wage penalty of 1.9%. However, overeducated male workers in rural areas earn a relatively high wage premium. Their wages are 10.3% higher than matched workers at the same education level.

This study simultaneously proves the relevance of human capital theory and job assignment theory. Based on the VV model, overeducated and undereducated workers generally receive a wage premium. However, if estimated based on the fields of study, a wage penalty for overeducated workers exists in seven fields of study: social sciences, journalism and information (ISCED Code 03), business sciences, administration and law (ISCED Code 04), information sciences and communication technologies (ICTs) (ISCED Code 06), engineering, manufacturing and construction sciences (ISCED 07), and agricultural, forestry, fisheries and veterinary sciences (ISCED 08). This wage penalty shows the unbalanced supply and demand for labour in those fields of study.

Even though overeducated workers still have economic benefits, this shows low productivity and losses employers bear. The high vertical educational mismatch is a form of external educational inefficiency that can cause Indonesia's low total factor productivity. Therefore, structured efforts must be made to reduce this vertical educational mismatch. Thus, the Indonesian government can refer to this study as input to developing strategies for enhancing human resources and education.

This study has a limitation. It did not examine the wage effect of horizontal educational mismatch. Therefore, we could not explain the differences in wages of vertical education mismatch workers based on horizontal mismatch in each field of study. Overeducated or undereducated workers can get premium or penalty wages due to horizontal mismatch. For example, a graduate majoring in teaching could get a higher wage when he becomes a technician and associate professional (overeducated and horizontally mismatched) in a relatively large company. Therefore, future studies are expected to fill this deficiency.

Acknowledgments

The authors would like to thank the Central Statistics Agency (the BPS) of the Republic of Indonesia for providing access to the National Labour Force Survey data.

References

Abidin, Z. Z., & Zakariya, Z. (2018). The impacts of over-education on economic growth in Malaysia. *Journal of Islamic, Social, Economics and Development*, *3*(11), 85–95.

Bauer, T. K. (2002). Educational mismatch and wages: A panel analysis. *Economics of Education Review*, *21*(3), 221–229.

Becker, G. S. (1992). Human capital and the economy. *Proceedings of the American Philosophical Society*, *136*(1), 85–92. Retrieved from http://www.jstor.org/stable/986801

Bedir, N. (2014). *The impact of over-education and under-education on earnings : Egypt in a post-revolutionary era*. School of Economics and Management, Lund University, Sweden.

Belfield, C. (2010). Over-education: What influence does the workplace have?. *Economics of Education Review*, *29*(2), 236–245.

Blázquez, M., & Budría, S. (2012). Overeducation dynamics and personality. *Education Economics*, *20*(3), 260–283.

Büchel, F., & Battu, H. (2003). The theory of differential overqualification: Does it work? *Scottish Journal of Political Economy*, *50*(1), 1–16.

Carmichael, F., Darko, C., & Kanji, S. (2021). Wage effects of educational mismatch and job search in Ghana and Kenya. *Education Economics*, *29*(4), 359–378.

Caroleo, F. E., & Pastore, F. (2018). Overeducation at a glance. Determinants and wage effects of the educational mismatch based on AlmaLaurea data. *Social Indicators Research*, *137*(3), 999–1032.

Carroll, D., & Tani, M. (2015). Job search as a determinant of graduate over-education: evidence from Australia. *Education Economics*, *23*(5), 631–644.

Chaya, P., Mashindano, O., & Chisomi, D. (2013). Over education and under education in the Tanzanian formal sector: determinants and policy implications. *International Journal of Education Economics and Development*, *4*(4), 371–391.

Chevalier, A., & Lindley, J. (2009). Overeducation and the skills of UK graduates. *Journal of the Royal Statistical Society*, *170*(2), 307–337.

Chung, Y. (2001). Overeducation and undereducation and their effect on earnings: Evidence from Hong Kong, 1986-1996. *Pacific Economic Review*, *6*(3), 401–418.

Clark, B., Joubert, C., & Maurel, A. (2017). The career prospects of overeducated Americans. *IZA Journal of Labor Economics*, *6*(3), 1–29.

Cohn, E., & Khan, S. P. (1995). The wage effects of overschooling revisited. *Labour Economics*, *2*(1), 67–76.

Croce, G., & Ghignoni, E. (2012). Demand and supply of skilled labour and overeducation in Europe: A country-level analysis. *Comparative Economic Studies*, *54*(2), 413–439.

Cutillo, A., & Pietro, G. Di. (2006). The effects of overeducation on wages in Italy: A bivariate selectivity approach. *International Journal of Manpower*, *27*(2), 143–168.

Devillanova, C. (2013). Over-education and spatial flexibility: New evidence from Italian survey data. *Papers in Regional Science*, *92*(3), 445–464.

Diem, A. (2015). Overeducation among graduates from universities of applied sciences: Determinants and consequences. *Journal of Economic & Financial Studies*, *3*(02), 63–77.

Diem, A., & Wolter, S. C. (2014). Overeducation among Swiss university graduates: determinants and consequences. *Journal for Labour Market Research*, *47*(4), 313–328.

Dolton, P., & Vignoles, A. (2000). The incidence and effects of overeducation in the U.K. graduate labour market. *Economics of Education Review*, *19*(2), 179–198.

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

Ege, A. A., & Erdil, E. (2023). Determinants of overlapping mismatch in the Turkish labour market. *Journal for Labour Market Research*, *57*(1), 1–9.

Frank, R. H. (1978). Why women earn less: The theory and estimation of differential overqualification. *American Economic Review*, *68*(3), 360–373. http://www.jstor.org/stable/1805267

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

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

Grunau, P., & Pecoraro, M. (2017). Educational mismatch and promotions to managerial positions: A test of the career mobility theory. *Applied Economics*, *49*(12), 1226–1240.

Gujarati, D. (2015). *Econometrics by example* (2nd ed). United Kingdom: Palgrave Macmillan Education.

Haddad, G. R. K., & Habibi, N. (2017). Vertical skill mismatch and wage consequences in low-skilled jobs: Evidence from Iran. *International Labour Review*, *156*(1), 45–72.

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.

Hout, M. (1984). Status, autonomy, and training in occupational mobility. *American Journal of Sociology*, *89*(6), 1379–1409.

Iriondo, I., & Amaral, T. P. (2016). The effect of educational mismatch on wages in Europe. *Journal of Policy Modeling*, *38*(2), 304–323.

Johnes, G. (2019). The incidence of and returns to ‘overeducation’: PIAAC evidence on the G7. *Minerva*, *57*(1), 85–107.

Jovanovic, B. (1979). Job matching and the theory of turnover. *Journal of Political Economy*, *87*(5, Part 1), 972–990.

Kain, J. F. (1992). The spatial mismatch hypothesis: Three decades later. *Housing Policy Debate*, *3*(2), 371–460.

Karakaya, G., Plasman, R., & Rycx, F. (2007). Overeducation on the Belgian labour market: Evaluation and analysis of the explanatory factors through two types of approaches. *Compare*, *37*(4), 513–532.

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

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

Kupets, O. (2015). *Skill mismatch and overeducation in transition economies*. IZA World of Labor. Retrieved from https://doi.org/10.15185/izawol.224.

Leuven, E., & Oosterbeek, H. (2011). Overeducation and mismatch in the labor market. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the Economics of Education* (pp. 283–326). Amsterdam, Netherlands : Elsevier B.V.

Li, I. W., Harris, M., & Sloane, P. J. (2018). Vertical, horizontal and residual skills mismatch in the Australian graduate labour market. *Economic Record*, *94*(306), 301–315.

Li, I. W., & Miller, P. W. (2015). Overeducation and earnings in the Australian graduate labour market : An application of the Vahey model. *Education Economics*, *23*(1), 63–83.

McFadden, D. (2021). Quantitative methods for analysing travel behaviour of individuals: Some recent developments. In D. A. Hensher & P. R. Stopher (Eds.), *Behavioural travel modelling* (pp. 279–318). London, England: Routledge.

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

McGuinness, S., Pouliakas, K., & Redmond, P. (2018). Skills mismatch: Concepts, measurement and policy approaches. *Journal of Economic Surveys*, *32*(4), 985–1015.

Mincer, J. A. (1974). *Schooling, experience, and earnings*. Cambridge, England: National Bureau of Economic Research.

Mugijayani, W. (2020). *Return to education and education mismatch in Indonesia*. (Doctoral Dissertation, Cardiff University, Cardiff, Wales). Retrieved from http://orca.cf.ac.uk/131497/.

Munsech, Q. C. (2019). Measuring overeducation: Incidence, correlation and overlaps across indicators and countries. *Social Indicators Research*, *145*(1), 279–301.

Neycheva, M. (2020). Qualification mismatch and growth in the Eastearn European region. *Balkan and Near Eastern Journal of Social Sciences*, *06*(04), 71–77.

Ordine, P., & Rose, G. (2017). Too many graduates? A matching theory of educational mismatch. *Journal of Human Capital*, *11*(4), 423–446.

Park, K., & Jang, D. (2017). The wage effects of over-education among young stem graduates. *Singapore Economic Review*, *63*(1), 1–20.

Pholphirul, P. (2017). Educational mismatches and labor market outcomes: Evidence from both vertical and horizontal mismatches in Thailand. *Education + Training*, *59*(5), 534–546.

Piracha, M., Tani, M., & Vadean, F. (2012). Immigrant over- and under-education : The role of home country labour market experience. *IZA Journal of Migration*, *1*(3), 1–21.

Quinn, M. A., & Rubb, S. (2006). Mexico's labor market: The importance of education-occupation matching on wages and productivity in developing countries. *Economics of Education Review*, *25*(2), 147–156.

Robert, P. (2014). Job mismatch in early career of graduates under post-communism. *International Journal of Manpower*, *35*(4), 500–513.

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

Rumberger, R. W. (1981). The rising incidence of overeducation in the U.S. Labor market. *Economics of Education Review*, *1*(3), 293–314.

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

Sattinger, M., & Hartog, J. (2013). Nash bargaining and the wage consequences of educational mismatches. *Labour Economics*, *23*, 50–56.

Schweri, J., Eymann, A., & Aepli, M. (2020). Horizontal mismatch and vocational education. *Applied Economics*, *52*(32), 3464–3478.

Senkrua, A. (2015). The mismatch in Thai labor market: Overeducation. *Journal of Economics*, *19*(1), 92-116. Retrieved from https://so01.tci-thaijo.org/index.php/CMJE/article/view/61018

Serikbayeva, B., & Abdulla, K. (2022). Education-job mismatch: implications for individual earnings and aggregate output. *Social Indicators Research*, *163*(2), 723–

752.

Sicherman, N. (1990). Education and occupational mobility. *Economics of Education Review*, *9*(2), 163–179.

Sicherman, N., & Galor, O. (1990). A theory of career mobility. *Journal of Political Economy*, *98*(1), 169–192.

Sitorus, F. M., & Wicaksono, P. (2020). The determinant of educational mismatch and its correlation to wages. *Jurnal Ekonomi Pembangunan*, *18*(2), 163–176.

Spence, M. (1973). Job market signaling. In P. Diamond & M. Rothschild (Eds.), *Uncertainty in Economics* (pp. 355–374). Amsterdam, Netherlands : Elsevier B.V. Retrieved from http://dx.doi.org/10.1016/B978-0-12-214850-7.50025-5.

Sun, H., & Kim, G. (2022). The wage effects of overeducation across overall wage distribution on university graduates: incidence, heterogeneity and comparison. *International Journal of Manpower*, *43*(5), 1144–1165.

Thurow, L. C. (1975). *Generating inequality*. London, England : The Macmillan Press Ltd.

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

Vahey, S. P. (2000). The great Canadian training robbery: evidence on the returns to educational mismatch. *Economics of Education Review*, *19*(2), 219–227.

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

Verhaest, D., & Omey, E. (2006). The impact of overeducation and its measurement. *Social Indicators Research*, *77*(3), 419–448.

Verhaest, D., & Omey, E. (2010). The determinants of overeducation: different measures, different outcomes? *International Journal of Manpower*, *31*(6), 608–625.

Verhaest, D., & Omey, E. (2012). Overeducation, undereducation and earnings: Further evidence on the importance of ability and measurement error bias. *Journal of Labor Research*, *33*(1), 76–90.

Wen, L., & Maani, S. A. (2018). A panel study of immigrants' overeducation and earnings in Australia. *International Migration*, *56*(2), 177–200.

Wu, N., & Wang, Q. (2018). Wage penalty of overeducation: New micro-evidence from China. *China Economic Review*, *50*, 206–217.

Wulandari, H., & Damayanti, A. (2021). Qualification mismatch dan upah di Indonesia [Qualification mismatch and wages in Indonesia]. *Jurnal Ekonomi Dan Kebijakan Publik Indonesia*, *8*(1), 45–57.

Zakariya, Z. (2014). The effects of over, required and under-education on earnings in manufacturing sector in Malaysia. *International Journal of Management Studies*, *21*(1), 83–109.

Zakariya, Z., & Noor, M. A. M. (2014). Workplace characteristics and determinants of over-education and overskilling in the manufacturing sector in Malaysia. *Jurnal Pengurusan*, *40*, 125–136.

Zheng, Y., Zhang, X., & Zhu, Y. (2021). Overeducation, major mismatch, and return to higher education tiers: Evidence from novel data source of a major online recruitment platform in China. *China Economic Review*, *66*, 1–16.

Appendix

Table 11: ISCED Code Descriptions (ISCED-F 2015 Version)

ISCED Code	Descriptions
01	**Education**
011	Education
018	Inter-disciplinary programs and qualifications involving education
02	Arts and Humanities
021	Art
022	Humanities (except languages)
023	Languages
028	Inter-disciplinary programs involving arts and humanities
03	**Social sciences, journalism, and information**
030	Social sciences, journalism, and information are not further defined.
032	Journalism and information
038	Inter-disciplinary programs and qualifications involving social sciences, journalism, and information
04	**Business, administration, and law**
041	Business and Administration
042	Law
048	Inter-disciplinary programs and qualifications involving business, administration, and law
05	**Natural sciences, mathematics, and statistics**
050	Natural sciences, mathematics, and statistics are not further defined
051	Biological and related sciences
052	Environment
053	Physical sciences
054	Mathematics and statistics
06	**Information and Communication Technologies (ICTs)**
061	Information and Communication Technologies (ICTs)
068	Inter-disciplinary programs and qualifications involving Information and Communication Technologies (ICTs)
07	**Engineering, manufacturing, and construction**
070	Engineering, manufacturing, and construction are not further defined
071	Engineering and engineering trades
072	Manufacturing and processing
073	Architecture and construction
078	Inter-disciplinary programs and qualifications involving engineering, manufacturing, and construction
079	Engineering, manufacturing, and construction not elsewhere classified
08	**Agriculture, forestry, fisheries and veterinary**
080	Agriculture, forestry, fisheries, and veterinary are not further defined
081	Agriculture
082	Forestry
083	Fisheries
084	Veterinary
088	Agriculture, forestry, fisheries, and veterinary not elsewhere classified

09	**Health and Welfare**
090	Health and welfare are not further defined
091	Health
098	Inter-disciplinary programs and qualifications involving health and welfare
10	**Services**
101	Personal services
102	Hygiene and occupational health services
103	Security services
104	Transport services

Source: United Nations Educational, Scientific and Cultural Organization (UNESCO)

Table 12: First Digit of ISIC Code Descriptions Rev 4, 2008

Code	Description
A	Agriculture, forestry, and fishing
B	Mining and quarrying
C	Manufacturing
D	Electricity, gas, steam, and air conditioning supply
E	Water supply; sewerage, waste management, and remediation activities
F	Construction
G	Wholesale and retail trade; repair of motor vehicles and motorcycles
H	Transportation and storage
I	Accommodation and food service activities
J	Information and communication
K	Financial and insurance activities
L	Real estate activities
M	Professional, scientific, and technical activities
N	Administrative and support service activities
O	Public administration and defense; compulsory social security
P	Education
Q	Human health and social work activities
R	Arts, entertainment, and recreation
S	Other service activities
T	Activities of households as employers; undifferentiated goods- and services-producing activities of households for own use
U	Activities of extraterritorial organizations and bodies

Source: United Nations Statistics Division (UNSD)

Glossary

[1]**ORU** stands for Overeducation, Required Education, and Undereducation. It is an economic model by Duncan and Hoffman (1981) that estimates the return to overeducation, required education, and undereducation

[2]**VV** stands for Verdugo and Verdugo. It is the economic model that was developed by Verdugo and Verdugo (1989) to estimate the wage penalty or premiums for vertically mismatched workers (i.e., overeducated or undereducated)

[3]**JA** stands for Job Analysis. It is a method to estimate the vertical educational mismatch by determining the years of required education by matching individuals' years of education attainment with occupational codes in ISCO

[4]**ISCO** stands for International Standard Classification of Occupation. It is a long list of occupational codes published by the International Labour Organization (ILO).

[5]**RM** stands for Realised Matched. It is a method to estimate the vertical mismatch by determining the value of required education through statistical values such as mean, median, mode, or range resulting from the distribution of educational levels of a sample group in each occupational group

[6]**CMT** stands for Career Mobility Theory. It is a theory by Sicherman and Galor (1990), which explains that one of the returns to education is occupational mobility within or across firms. Increasing the level of education will increase occupational or career advancement so that highly educated workers who do not get promotions are likely to leave their old workplaces

[7]**JCT** stands for Job Competition Theory. It is a theory developed by Thurow (1975) to explain the phenomenon of job competition, job distribution, and wage determination based on job competition

[8]**ISCED** stands for International Standard Classification of Education. It is a list of educational fields and programs

[9]**SOE** stands for State-Owned Enterprises. It is the list of companies whose shares are majority-owned by the government

[10]**DOVER** stands for Dummy Overeducation. It is a categorical overeducation variable, valued at 0 if not overeducated and 1 if overeducated

[11]**DUNDER** stands for Dummy Undereducation. It is a categorical undereducation variable, valued at 0 if not undereducated and 1 if undereducated

[12]**YOVER** stands for Years of Overeducation or surplus years of schooling. It is an overeducation variable with a value of more years of education. YOVER is obtained from YEDUC - YREQ if YEDUC > YREQ

[13]**YREQ** stands for Years of Required Education. It is a variable that shows the years required to obtain or work in an occupation. YREQ can be estimated using 2 approaches: subjective and objective. The subjective approach asks respondents directly or indirectly about the minimum education needed to obtain or work on a job. At the same time, the objective is divided into 2 methods, namely JA and RM

[14]**YUNDER** stands for Years of Undereducation or deficit years of schooling. It is an undereducation variable with the value of years of education deficit. YUNDER is obtained from YREQ – YEDUC if YREQ > YEDUC

[15]**YEDUC** is Years of Education attainment or education completed by individuals. We estimate the value of YEDUC by converting the sample education classification data into years. The sample with secondary school = 12 years, diploma 1 = 13, diploma 2 = 14, diploma 3 = 15, diploma 4 or bachelor = 16, magister = 18, and doctoral = 22

[16]**OLS** stands for Ordinary Least Squares. It is one of the regression estimation methods with the smallest ordinary square method

[17]**GOF** stands for Goodness of Fit. It is a statistical measure that shows the fit of the econometric model built by examining the simultaneous influence of explanatory variables on the regressor variables