I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17161

Overeducation, Earnings and Job Satisfaction among Graduates in China

Melanie Jones
Ezgi Kaya
Jiarui Nan

JULY 2024

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17161

Overeducation, Earnings and Job Satisfaction among Graduates in China

Melanie Jones
Cardiff University and IZA

Ezgi Kaya
Cardiff University and GLO

Jiarui Nan
University of Sheffield

JULY 2024

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Overeducation, Earnings and Job Satisfaction among Graduates in China*

Exploiting rich nationally representative longitudinal data from the China Family Panel Studies this paper explores the relationship between overeducation, earnings and job satisfaction among graduates in China. We find consistent evidence, across multiple measures of overeducation, of wage and job satisfaction penalties that are not explained by personal and work-related characteristics. Despite attention within the literature, we find a modest role for differences in academic subject and, cognitive and non-cognitive skills as drivers of these penalties. In contrast, controlling for time-invariant unobserved heterogeneity reduces the size and, in many cases, removes the statistical significance of overeducation penalties, aligned to the importance of other unobserved individual heterogeneity.

JEL Classification: J24, J28, J31

Keywords: overeducation, China, earnings, job satisfaction, cognitive and non-cognitive skills, unobserved heterogeneity

Corresponding author:
Jiarui Nan
Department of Economics
University of Sheffield
Sheffield, S1 4DT
United Kingdom
E-mail: j.nan@sheffield.ac.uk

* The data are from China Family Panel Studies (CFPS), funded by Peking University and the National Natural Science Foundation of China. The CFPS is maintained by the Institute of Social Science Survey of Peking University. Jiarui Nan is grateful to the Julian Hodge Foundation for a PhD scholarship on which this analysis is based. We are also grateful to Keith Bender and Andrew Henley, as well as participants at the 2023 European Association of Labour Economists Conference in Prague for comments on an earlier version of this analysis.

1. Introduction

Interest in the implications of graduate overeducation, that is, the possession of degree level qualifications which are not required in an individual's current job, intensified in developed countries with the expansion of higher education (see McGuinness, 2006 for a review). It is perhaps unsurprising since overeducated workers are, by definition, not fully utilising their education relative to those whose education matches their job requirement, that there is widespread evidence that overeducated workers face a wage penalty (for a meta-analysis see Groot and van den Brink, 2000) and lower job satisfaction relative to those who are not overeducated (see, for example, Green and Zhu, 2010). The expansion of education has, however, since been a feature in many developing countries, extending concern that the inefficiencies and personal costs associated with graduate overeducation might have also spread internationally (see Mehta *et al*., 2011; Zheng *et al.,* 2021 among others). Despite this, nationally representative evidence on the prevalence and implications of overeducation in developing countries remains scarce (Quinn and Rubb, 2006; Zheng *et al.,* 2021), at least partially due to data being more restricted (see Wu and Wang, 2018; Zheng *et al.,* 2021).[1]

In this paper, we utilise data from a large nationally representative longitudinal survey, the China Family Panel Studies (CFPS), to provide new evidence for China, where dramatic growth in the proportion of young people attending higher education since the late 1990's, albeit coinciding with significant economic development, has the potential to increase the prevalence and importance of graduate overeducation. In particular, we focus on graduates and explore the relationship between individual overeducation, pay and job satisfaction. Despite receiving less attention, the latter captures additional dimensions of

[1] For evidence for developing countries outside China see Quinn and Rubb (2006) for Mexico, Mehta *et al*. (2011) for Mexico, Thailand, the Philippines and India and Aslam *et al*. (2012) for Pakistan.

job quality (such as autonomy, security and the work environment) central to employee wellbeing and behaviour (such as job performance, engagement and turnover). Importantly, the comprehensiveness of the CFPS allows us to address two key concerns in the international literature (McGuinness, 2006) that have yet to receive the same attention in developing countries. First, we utilise multiple measures of overeducation to explore the robustness of patterns across objective, subjective, and statistical measures of overeducation. Second, we explore the extent to which wage and job satisfaction penalties associated with graduate overeducation can be attributed to variables that are typically unobserved. In particular we address concerns that overeducated workers might differ in the nature of their education, possess lower skill levels or underutilise their skills relative to their non-overeducated counterparts.[2] We further consider skill along multiple dimensions, including non-cognitive dimensions typically neglected in the literature but well established as important for labour market outcomes (see, for example, Heckman *et al*., 2006). Finally, we utilise the longitudinal element of the CFPS and estimate individual fixed effects models thereby capturing all time-invariant unobserved heterogeneity which might otherwise bias our estimates (see, for example, Tsai, 2010).

We document the scale of overeducation among Chinese graduates, which varies from 37% to 47%, depending on the measure, and provide evidence of robust wage and job satisfaction penalties which exists even after controlling for personal and work-related characteristics. Despite attention within the international literature our evidence suggests subject, cognitive and non-cognitive skills, and cognitive skill mismatch play a modest role in explaining these penalties. In contrast, we find a more important role for individual-specific fixed effects capturing time-invariant unobserved heterogeneity, with a general

[2] In doing so we explore whether the penalty to overeducation reflects lower skills among overeducated workers, which themselves reduce wages or job satisfaction, or that overeducation coincides with being overskilled and the overeducation penalty therefore reflects a penalty to skill mismatch.

dampening of the penalties, many of which become statistically insignificant. The latter is consistent with the presence of potentially neglected individual unobservable characteristics which explain the relationship between overeducation and earnings and job satisfaction.

Our analysis contributes comprehensive and nationally representative evidence to the emerging literature on China which is often restricted in scope, including being focused entirely on pay (Wu and Wang, 2018; Liu *et al.*, 2021; Zheng *et al.*, 2021), limited to specific measures of overeducation (Wu and Wang, 2018; Liu *et al*., 2021; Zheng *et al.*, 2021) or in geographic coverage (for example, Wu and Wang, 2018), based on small cross-sectional samples (Wu and Wang, 2018; Liu *et al*., 2021) and/or fails to consider important omitted variables such as skills (Liu *et al*., 2021; Zheng *et al.*, 2021).

As such, our evidence adds new insights from a developing country context to the international literature, particularly on the role of measurement of overeducation and unobserved skills in explaining pay and job satisfaction penalties (see, for example, Chevalier, 2003; Chevalier and Lindley, 2009; Green and Zhu, 2010; Levels *et al.*, 2014). The key role played by the choice between including direct skill measures and controlling for time-invariant unobserved heterogeneity aligns with findings in the existing literature for developed countries.[3] Indeed, overall, and in contrast to prior evidence for developing countries which is often sensitive to the country context (Mehta *et al*., 2011), our findings for China suggest that graduate overeducation shares many of the features of developed countries.

The remainder of this paper is structured as follows. Section 2 provides an overview of the education and labour market context in China. Section 3 introduces the CFPS data

[3] The absence of a relationship between overeducation and labour market outcomes after accounting for unobserved time-invariant individual heterogeneity is observed in the US (Tsai, 2010), Australia (Mavromaras *et al*., 2013), Germany (Bauer, 2002) and Canada (Frenette, 2004).

and measures employed in the analysis. In Section 4 we set out our empirical approach and present our results. Section 5 briefly concludes.

2. Education and overeducation in China

Our focus is on post-compulsory tertiary level education (college and above) in China. This comprises graduates from three-year colleges of professional training and higher vocational education, leading to a vocational college qualification, and universities, typically offering four or five-year programs leading to a bachelor's degree. Admission is through a centralized system based on scores in the standardized National College Entrance Examinations, leading to tiered placements.[4] Graduates from universities also have the opportunity to pursue postgraduate education, such as master's or PhD degrees.

The tertiary education in China has undergone dramatic changes since the late 1970s "reform and opening up" policy. During this period, China has experienced significant industrialisation, with strong economic growth and a shift from public to private sector employment (see, for a review, Liu, 2020). This has also been a period of urbanisation and stark regional inequality, with higher educational attainment and hourly earnings in urban compared to rural areas reinforced by a household registration system (i.e., 'hukou') (see Yin, 2004; Wang *et al*., 2014).

With economic reform and market opening, China's education system also underwent significant expansion and transformation to satisfy the demand for graduates (Wu and Zheng, 2008).[5] The higher education expansion policy "Action Plan for Revitalizing Education in the 21st Century" was proposed by the Chinese Ministry of

[4] Colleges and universities in China are classified into three tiers in descending order of prestige and entry requirements: Key Universities, which are mostly research-intensive institutions; Ordinary Universities; and vocational training colleges (see, for further discussion, Kang *et al*., 2021). For a theoretical analysis of the Chinese college admission system see Chen and Kesten (2017).

[5] Wu and Zheng (2008) suggest that catering to parents seeking increased higher education opportunities for their children, alleviating urban unemployment, and promoting the development and utilisation of China's human resources are among the other arguments for China's higher education expansion in 1999.

Education in 1999, where the original goal was to increase the gross enrolment rate of tertiary education to 15% by 2010. In fact, the growth in enrolment far exceeded this and was 25.3% by 2010 and 72.0% by 2022 (see World Bank, 2023).[6]

Further to the expansion policy, from the 1990s to the 2000s, the Chinese government gradually reduced subsidies for higher education, which had been heavily subsidised under the earlier central planning regime. Some argue that this resulted in severe financial constraints for low-income families, and that not everyone benefited from the expansion of higher education (Li *et al.*, 2014). Another significant higher education reform in the 1990s abolished the centralised planning system for allocating college graduates to work positions, reducing government intervention and increasing employment pressures for college graduates following the expansion (Freeman, 2010; Li *et al*., 2014). Empirical evidence suggests that the expansion of higher education in China led to an increase in the unemployment rate among young college graduates from 2000 to 2005 by 6 to 9%, consistent with short-term pressure (Li *et al.*, 2014). However, subsequent analysis indicates that this impact did not persist in the medium run, as no consistently significant effect has been found on the unemployment of university graduates in 2010 (Xing *et al*., 2018).

Despite the increase in supply, evidence suggests that the return to tertiary education has not experienced a dramatic decline following the expansion policy. Focusing on urban male workers in China, Meng *et al*. (2013) show that the most significant increase in the return to tertiary education occurred during the 1990s, and by the end of that decade, the tertiary education premium is found to reach 30 to 50% (see, for a review, Li *et al*., 2014). Although the trends became flatter following the higher education policy, the

[6] These figures surpass the East Asia and the Pacific average of 25.6% in 2010 and 58.1% in 2022 but are lower than the OECD average of 65.6% in 2010 and 78.7% in 2022 (see World Bank, 2023).

estimates by Meng *et al*. (2013) suggest that tertiary education premium was still around 50% in the early 2000s. More recently, however, Asadullah and Xiao (2020) find a decline in the return to bachelor education and above, from 36.5% in 2010 to 28.9% in 2015. However, Kang *et al*. (2021) show that the decline varies based on the subject and institutional quality, with reduced returns observed for all graduates except those from more prestigious key universities studying subjects such as law, economics, and management (LEM), or sciences, technology, engineering, and math (STEM).

A separate body of Chinese literature also highlights the increase in the incidence of overeducation among graduates in the labour market, attributed to the rise in the supply side resulting from the expansion of higher education although this is arguably considered as a temporary phenomenon (Li *et al*., 2008). More recently, however, Zheng *et al.* (2021), analysing textual data from a large online recruitment platform in China by matching job seekers' qualifications from their resumes with recruiters' requirements from job postings published between June and August 2017, find that half of the online jobseekers on the platform, of whom almost 90% had a college degree or higher, were overeducated.

Consistent with international literature (see, for a review, McGuinness, 2006), the evidence for China also suggests that overeducated graduates experience a wage penalty in the labour market. For instance, using data from the 2012 World Bank Skills Towards Employability and Productivity (STEP) Survey in China and measuring overeducation based on individuals' self-assessment of their education level relative to that required in their job, Wu and Wang (2018) find that overeducated workers with a tertiary education level experience a 20% loss in hourly wage compared to matched individuals. Consistent with this, the textual analysis of Zheng *et al.* (2021) also reveals a 5.1% overeducation penalty in monthly pay, although this varies depending on subjects, institutional quality, as well as across industries and cities.

A number of Chinese studies also explore the relationship between overeducation and job satisfaction, although evidence specifically for graduates is limited. One exception, Shi *et al*. (2020), using data from the 2020 Panel Study of Chinese University Students explore the impact of overeducation on new graduates' job satisfaction by gender. Their findings indicate mixed results based on the overeducation measure employed. While the objective measure of overeducation is found to be positively correlated with job satisfaction for females and shows no significant correlation for males, the subjective measure of overeducation is consistently found to be negatively correlated with job satisfaction regardless of gender. In fact, the latter is consistent with the findings of Ma *et al*. (2022) for all workers, rather than just graduates, based on CFPS data from 2014 and 2018, which show a negative relationship between the subjective measure of overeducation and job satisfaction. Ma *et al*. (2022) further show that this negative relationship is partly mediated by the wage penalty associated with overeducation.

3. CFPS data

The analysis utilises data from CFPS, a large nationally representative longitudinal survey in China designed by Peking University and containing detailed information on education and labour market outcomes.[7] The survey has been administered biannually since 2010, has a target sample of 16,000 households (or 50,000 individuals) covering the majority of Chinese provinces.[8] We pool data between 2014-2020, the period over which information on our key variables, including earnings, is consistent.[9] Our focus is on graduates defined as those who have completed tertiary education (including college, undergraduate university and postgraduate education). The CFPS has previously been used to explore the

[7] For further information about the CFPS data, see the official website at https://www.isss.pku.edu.cn/cfps/en.
[8] In the initial 2010 survey, 25 of the 34 Chinese provinces were sampled.
[9] Throughout we present estimates from models using the pooled data since we find limited variation in the pay or job satisfaction penalties associated with overeducation over the period (results available upon request).

rate of return to education in China (Kang *et al*., 2021) and collects information aligned to established measures in this literature. It has several advantages in this context, particularly the ability to derive three established measures of overeducation, information on both earnings and job satisfaction, the availability of information on academic subject and cognitive and non-cognitive skills and, its longitudinal nature which allows us to control for individual-specific fixed effects capturing time-invariant unobserved heterogeneity.[10]

Our analysis sample is restricted to prime age employees (aged 25-54) who are not currently in education and who are wage earners.[11] After excluding those with missing values for any of the variables used in the analysis, our maximum unbalanced sample contains 6,170 observations from 3,201 employees.[12]

Overeducation

In contrast to most of the literature, the CFPS includes information from which to derive three established measures of overeducation, namely, objective, statistical and subjective overeducation. [13] The objective measure relies on assessments relative to external benchmarks such as those agreed by job analysts and is constructed based on the International Standard Classification of Occupations (ISCO), which are linked to CFPS occupational codes (defined according to the national standard occupational classification of the People's Republic of China at the three-digit level). Following Romero and Jimenez (2017), the first three one-digit ISCO occupations are defined as matched for graduates, the remaining occupations (one-digit ISCO codes 4-9) are classed as overeducated.

[10] Other potential surveys, such as the STEP survey, the China Health and Nutrition Survey (CHNS) and the China General Social Survey (CGSS) are limited in one or more respects, including having a restricted geographic coverage, the absence of longitudinal data and/or less comprehensive information on education/skills.

[11] As a result, all self-employed workers and employees who do not report pay are excluded from the analysis.

[12] Further details of the panel properties of the data are available in Online Appendix Table A1.

[13] There is no consensus in the literature as to the optimal measure, with each having been subject to its own criticisms (see, for example, Hartog, 2000; Groot and van den Brink, 2000 and McGuinness, 2006) and choices typically determined by data availability.

The statistical measure (sometimes referred to as a realised match approach) is calculated based on a comparison between own education in years and the mode years of education of all employees within the two-digit occupation in each year.[14] We define overeducation as where the years of education (imputed based on the level of highest education reported) exceed the occupation mode. All other employees are defined as not overeducated.[15]

The subjective measure requires individuals to self-assess their education level relative to that required in their job. Here we base this on a question which asks workers to reflect on the education level that is required to be qualified to do their job, with responses ranging from no schooling to doctoral level.[16] Responses below tertiary level form overeducation, with responses at tertiary level classed as non-overeducated. In a similar manner to Kler (2005) for Australia, Verhaest and Omey (2006) for Belgium and Mateos-Romero and Salinas-Jiménez (2017) for Spain we compare our findings across our overeducation measures to explore the sensitivity of our findings to the measurement issues involved.

According to Table 1, the percentage of the sample who are overeducated is 36.9%, 46.6% and 46.1% for the objective, statistical and subjective measures, respectively.[17] Our estimates are aligned to prior evidence from China, where about 50% of graduates are

[14] See Battu and Sloane (2004) for a discussion of the advantages of using the mode relative to the mean education level. While the CFPS contains more than 300 three-digit occupation codes the samples within each three-digit occupation are too small to provide reliable information. For the same reason we follow Battu and Sloane (2004) and aggregate some of the two-digit codes into larger related occupations such that we utilise 43 separate occupational groups, with a minimum of 10 observations in each. Further details are available upon request.

[15] For the subjective and statistical measures, we can distinguish between matched and undereducated workers but the sample of undereducated employees is small and so we combine with matched throughout and explore the sensitivity of our estimates in Online Appendix Tables A8 and A9.

[16] Since 2014 this question has only been asked to workers who change their job. We assume the response is time-invariant for those who remain in the same job, but the absence of information in 2014 for some employees results in a smaller sample for this measure.

[17] We explore the overlap between the three measures in Online Appendix Table A2. Consistent with existing international evidence, while there is considerable overlap, each measure contains a distinct element of coverage.

overeducated according to the subjective measure (Wu and Wang, 2018) and based on a measure of intended overeducation derived from online recruitment platforms (Zheng *et al.*, 2021).[18]

Hourly pay

In line with the established literature, our main dependent variable is (log) gross hourly pay, (Yuan RMB) which adjusts pay for hours worked. It is constructed based on an annualised measure derived from average usual monthly earnings and weekly working hours. It therefore includes subsidies, performance-related and overtime pay and payment in-kind.[19] Table 1 presents mean log hourly earnings by overeducation. The observed pay penalties for overeducation are -20.78%, -4.40% and -20.47% for the objective, statistical and subjective measures, respectively.[20]

Job satisfaction

In the CFPS, all employees are asked about their overall level of job satisfaction, as well as facets relating to pay, safety, environment, working hours and promotion. Overall satisfaction is measured by responses to the following question: '*In general, how satisfied are you with this job?'*. In each case responses vary from the lowest satisfaction level 1 (*totally dissatisfied*) to the highest satisfaction level 5 (*totally satisfied*).[21] On average job satisfaction is lower among overeducated relative to non-overeducated employees (Table 1), with the mean gap of 0.142, 0.140 and 0.145 points (about 4%), for the objective, statistical and subjective measures, respectively.

[Table 1 here]

[18] In Online Appendix Table A3 we report transitions in overeducation status. A greater prevalence of transitions is observed for the objective and statistical measures of overeducation relative to the subjective measure. Rates of movement out from overeducation exceed the rates of movement into overeducation for all three measures.

[19] We explore the sensitivity of our estimates to the exclusion of outliers in Online Appendix Table A8.

[20] Throughout percentages are calculated as *exp*(log points) – 1.

[21] While the response options have remained consistent across waves in the CFPS, unlike in 2014 and 2016, the neutral option was not explicitly stated by interviewers in 2018 and 2020.

Explanatory variables

The CFPS also contains detailed information on personal and work-related characteristics well-established to affect earnings and job satisfaction. Our controls for personal characteristics include age (and age squared), dummy variables which capture (male) gender, details about the nature of (graduate) education, distinguishing between (more vocational) college, university and postgraduate attainment, minority ethnicity, marital status (married), household registration status (non-agricultural hukou) and (30) provinces of residence. Work-related characteristics include a set of dummy variables to capture (formal) contract, sector and industry (20 one-digit categories).[22] Given its use in the construction of objective and statistical overeducation measures we exclude occupation in our benchmark specifications but explore the sensitivity of our findings to its inclusion (see Online Appendix Tables A8 and A9).

In additional specifications we control for a series of variables which are not routinely available in survey data and therefore enable us to explore the impact of potentially important omitted dimensions of education and skills. First, given subject of degree is well-established to influence earnings and employment opportunities (Robst, 2008) we further control for graduate subject, where subject is defined across the following groups: Economics, Law, Education, Literature, Science, Engineering, Agronomy, Medicine, Management, and Other (including Arts, Philosophy and History).[23,24]

[22] Information on job tenure is not collected for those who do not change job. Using longitudinal information restricts the sample due to attrition and so we do not include tenure in our benchmark specifications. The results are not sensitive to its inclusion (see Online Appendix Tables A8 and A9).

[23] From 2014, CFPS collects subject information only for those who change education level. Given our focus on graduates we assume subject is time invariant and utilise information from 2010 and 2012 but attrition restricts the sample size.

[24] Unfortunately, information on institutional quality (such as the distinction between Key and Ordinary Universities) which is a potential determinant of overeducation and other labour market outcomes (see Zheng *et al.*, 2021) is only available for a relatively small subsample and so is not included in our benchmark specifications. We explore its influence in Online Appendix Tables A8 and A9.

Differences in skills among individuals with the same level of education may arise from factors such as innate ability, parental background or school quality. These in turn can influence their productivity and earnings. We allow for skill heterogeneity following Sohn (2010), Levels *et al*. (2014) and Nieto and Ramos (2017) and introduce controls for cognitive and non-cognitive skills. As is typical, we measure cognitive skills using literacy and numeracy test scores, which are administered in face-to-face interviews, and where values range from 0 to 34.[25] Given the focus on skill mismatch (see Allen and Van der Velden, 2001), we also generate two measures of skill utilisation. Similar to the statistical measure of overeducation, being overskilled is determined by comparing the mean level of (1) literacy and (2) numeracy skill among employees within each two-digit occupation in the same year (subject to a minimum sample size of 10).[26] Specifically, being overskilled is defined as the respective skill level being one or more standard deviations greater than the corresponding occupation mean. The remaining employees are defined as not overskilled.

As is now widely recognized, individuals' labour market outcomes, including overeducation, are affected by non-cognitive as well as cognitive skills (Heckman *et al*., 2006; Blázquez and Budría, 2012). We explore the former using the standard big five personality traits of openness, conscientiousness, extroversion, agreeableness and neuroticism, plus a measure of locus of control, to capture channels such as motivation, perseverance and networks in job search (see Palczyńska, 2021). Each measure is scored on a scale from 1 to 5 which increases with the relevant dimension of personality/locus of

[25] See Liu (2023) for a previous application utilising these measures. Since these measures are not asked in every CFPS wave, we assume cognitive skills are time invariant and utilise the score when first measured during the panel. The sample size is restricted as a result.
[26] No subjective measures of skill utilisation are available in the CFPS.

control.[27] Therefore, our most comprehensive specification compares overeducated and non-overeducated workers with comparable cognitive and non-cognitive skills.

Online Appendix Table A5 contains a full set of summary statistics for the explanatory variables by overeducation. They identify several characteristics associated with graduate overeducation, including that overeducated employees are more likely to be male, less likely to hold non-agricultural hukou status, and are less likely to hold a formal contract or work in the public sector. In terms of skills, overeducated employees have slightly lower literacy and numeracy test scores but are considerably more likely to be overskilled, consistent with overeducation being more closely related to skill mismatch than skill levels. Overeducation, at least according to the objective and statistical measures has also declined over the period.

4. Overeducation, earnings and job satisfaction

4.1 Quantifying earnings and job satisfaction overeducation penalties

We quantify the relationship between overeducation and earnings and job satisfaction using the Verdugo and Verdugo (1989) approach widely adopted in the international literature (see, for example, Chevalier, 2003; Green and Zhu, 2010) where the relevant outcome (the natural logarithm of gross hourly earnings or job satisfaction rank), Y_{it}, for individual i in year t is regressed on a binary indicator of being overeducated (O_{it}) as defined by one of the three measures introduced above as follows:[28]

$$Y_{it} = \alpha + \pi_t + \gamma O_{it} + X_{it}\beta + Z_{it}\varphi + \varepsilon_{it}. \qquad (1)$$

[27] Each personality trait/locus of control measure reflects the average of relevant items (see Online Appendix Table A4). We assume these are time invariant and utilise the score when first measured during the panel. All measures are increasing in the respective non-cognitive skill. For neuroticism, the index is measured in increasing ability to deal with neurotic situations.

[28] We adopt the same approach for job satisfaction despite the ordered nature of the dependent variable to facilitate the inclusion of fixed effects. See Ferrer-i-Carbonell and Frijters (2004) among others for a similar approach.

Throughout, we include year fixed effects (π_t) and a constant (α), and ε_{it} is a random error term. We successively build up the specification, adding the personal (X_{it}) and work-related (Z_{it}) characteristics outlined above. In this way, we explore whether the return to overeducation (γ), measured relative to non-overeducated graduates, typically found to be a penalty ($\gamma < 0$), can be explained by differences in other observable characteristics. In further specifications we explore the impact of introducing additional controls for key sources of omitted variable bias identified in the literature including subject, as well as cognitive and non-cognitive skills. In a final specification we replace time-invariant individual characteristics with individual fixed effects to account for time-invariant individual heterogeneity, which in addition to ability and non-cognitive skills might reflect motivation or vocational skills. Throughout standard errors are clustered at the individual level.

We present the results for hourly earnings in Table 2 where our focus is on the coefficient estimate on overeducation ($\hat{\gamma}$). We present the estimates for each of the three measures of overeducation (Panels A-C, respectively) and present successively more comprehensive models (columns (1)-(3), respectively). [29] Our evidence suggests a consistent and significant earnings penalty to overeducation after controlling for personal characteristics (column (2)).[30] While the direction of the relationship is consistent across the alternative measures of overeducation the magnitude is largest for the subjective measure (22.4%) and smallest for the statistical measure (6.6%). The inclusion of work-related characteristics (column (3)) has a relatively limited impact on the coefficient estimates and leaves our main conclusions unchanged. In this respect our findings are

[29] The full set of coefficient estimates from the underlying regression models (in column (3)) is presented in Online Appendix Table A6.
[30] It is important to control for education level when modelling the statistical measure of overeducation since there is a direct relationship between education and overeducation.

consistent with previous studies for China such as Wu and Wang (2018) for Kunmin City (22%), Liu *et al*. (2021) for 25 local universities in China (6.6%) and Zheng *et al.* (2021) who use intended overeducation based on data from online recruitment platforms (5.1%).

[Table 2 here]

The corresponding estimates for job satisfaction are presented in Table 3, where again we observe evidence of a consistent penalty across the three measures of overeducation. Here the magnitude of the coefficient estimates is more similar across the measures, with overeducation reducing job satisfaction by between 0.12 and 0.15 points on the five-point scale (between 3-4%) (column (1)). The relationship is dampened by the inclusion of personal and work-related characteristics (in column (2) and column (3), respectively) but this is most pronounced for the statistical measure of overeducation, where the relationship halves (and is now only weakly significant). Nevertheless, aligning to previous evidence from developed countries (see McGuinness, 2006), we observe a consistent job satisfaction penalty for overeducation even after accounting for the observable characteristics of individuals and their jobs.[31]

[Table 3 here]

We explore the robustness of our findings for pay and job satisfaction based on our most comprehensive specification in Online Appendix Tables A8 and A9, respectively. For both measures we examine the role of excluding the oversampling of selected provinces as part of the survey design (column (1)), controlling for broad (one-digit) occupation categories (column (2)), controlling for firm size (column (3)) and job tenure (column (4)) respectively (each of which is only available for a subsample of observations),

[31] Of course, the penalties to job satisfaction might reflect wage penalties but further analysis (see Online Appendix Table A7) across the facets of job satisfaction suggests the relationship is not exclusively with satisfaction with pay but extends to safety, the work environment and promotion opportunities. Further, the relationship between overall job satisfaction and overeducation remains significant for the objective and subjective measures after controlling for hourly pay (see Online Appendix Table A9).

and for institutional quality by distinguishing Key Universities from Ordinary Universities for the subsample where it is available (column (5)). We further distinguish undereducation from being matched (column (6)) and restrict the analysis to full-time employees (defined, consistent with Contract Law, as working 24 hours per week) (column (7)). In column (8) we present estimates with a correction for selection into paid employment (versus non-employment or self-employment) following Heckman (1979) where proxies for caring responsibilities are used as exclusion restrictions. For pay we explore the exclusion of hourly pay outliers in (column (9)). For job satisfaction we additionally present a specification (column (9)) which controls for pay to explore whether the overeducation job satisfaction penalty exists over any impact on pay. With one exception they confirm the robustness of the main results. The significance of the job satisfaction penalty associated with statistical overeducation is sensitive to the model specification. Further, as might be expected given its (occupational) construction, the estimates for the pay penalty associated with the statistical measure of overeducation are smaller, and become insignificantly different from zero, after controlling for occupation (column (3)).

4.2 Understanding earnings and job satisfaction overeducation penalties

In Tables 4 and 5 we explore the extent to which our benchmark estimates relating to pay and job satisfaction penalties after adjusting for personal and work-related characteristics are sensitive to the inclusion of key omitted variables namely education subject (column (1)), cognitive skills (column (2)) and skill mismatch (column (3))), and non-cognitive skills (column (4)), which might otherwise explain overeducation penalties. In Table 4 we find that the pay penalty associated with the three measures of overeducation is generally robust to the inclusion of controls for subject, cognitive skills and cognitive skill mismatch, and non-cognitive skills. In this respect our analysis does not support any of these variables as key channels through which overeducation affects hourly earnings and confirms a

residual unexplained pay penalty to overeducation. In contrast to what is often assumed, we find limited evidence that the penalty to the objective and subjective measure overeducation reflects lower cognitive and non-cognitive skills, or cognitive skill mismatch among overeducated workers. However, these measures play a more important role for the penalty associated with the statistical measure of overeducation (Panel B), which is no longer statistically significant. While our evidence therefore suggests broader generalisability of the findings for Kunmin City in China based on the subjective measure (Wu and Wang, 2018), it also highlights heterogeneity in the influence of unobserved education and skills across overeducation measures.

In relation to job satisfaction, for the subjective measure of overeducation (Panel C) the job satisfaction penalty is robust in Table 5 and suggests the role of overeducation *per se* rather than any of these omitted variables.[32] The estimates do, however, differ depending on the particular measure employed. While the findings for the objective measure (Panel A) are robust to the inclusion of the individual measures of education/skills, the job satisfaction penalty becomes insignificant after their simultaneous inclusion (column (5)). For the statistical measure of overeducation (Panel B) the estimates vary across specifications, being smaller and in some cases statically insignificant, consistent with the importance of subject and skill mismatch in explaining the relationship between overeducation and job satisfaction.[33]

The conclusions both for pay and job satisfaction are, however, different if we focus on the coefficient estimates from the fixed effects model in column (6). Except for the job satisfaction penalty associated with the subjective measure of overeducation the

[32] This is despite the important relationship between non-cognitive skills (as measured by the big five personality traits and locus of control) and job satisfaction.
[33] We estimate our benchmark models on a common smaller subsample (from column (5)) and except for the relationship between statistical overeducation and job satisfaction (which is insignificantly different from zero) our findings remain the same confirming it is the influence of our measures of education/skill rather than the sample is driving the results (available upon request).

relationships are generally dampened and often statistically insignificant suggesting the relationship between overeducation and pay, and overeducation and job satisfaction might also reflect other forms of individual-specific time-invariant unobserved heterogeneity. Surprisingly, the job satisfaction penalty to subjective overeducation is strengthened (Table 5, Panel C). Of course, such estimates are identified by transitions in overeducation status which are not necessarily representative (Chevalier, 2003) but it is worth noting that the pattern of the findings for China is similar to that for developed countries (Bauer, 2002; Frenette, 2004; Tsai, 2010; Mavromaras *et al*., 2013).[34]

[Tables 4 and 5 here]

5. Conclusions

Using nationally representative longitudinal data from the CFPS 2014-2020 this paper explores the relationship between overeducation, earnings and job satisfaction among graduates in China where, despite rapid growth in educational attainment, the potential consequences of overeducation have been neglected. Depending on the precise measure employed we find that between 37% and 47% of graduates are overeducated, consistent with a significant prevalence of graduate overeducation in China and suggestive of education expansion which has not been matched by education utilisation among employers.

We find consistent evidence of wage and job satisfaction penalties across three established measures of overeducation, which are not explained by personal and work-related characteristics. Further, despite the focus in the literature, we find a modest role for differences in the subject of graduate education, cognitive skills and skill mismatch (despite this being considerably higher among overeducated workers), and non-cognitive

[34] De Santis *et al*. (2022) find reduced but significant wage penalties to graduate overeducation when applying fixed effects methods to data from Argentina.

skills as drivers of several of the overeducation penalties. The role of these factors does, however, vary between the alternative measures of overeducation, being a more important explanation for penalties associated with the statistical measure of overeducation.

In contrast, overeducation penalties are generally smaller in magnitude and commonly absent after controlling for individual fixed effects consistent with potential additional unobserved heterogeneity. In this respect the evidence for China is aligned with prior evidence based on developed countries, both in terms of the prevalence of overeducation and, the inability of direct skill measures and ability of panel data approaches to explain wage and job satisfaction penalties. This is in contrast with prior evidence from some developing countries where even the existence of overeducation is debated (Mehta *et al*., 2011). Of course, the modest role for skills in some specifications might reflect that the measures available in the CFPS, while capturing cognitive and non-cognitive dimensions, are necessarily incomplete.[35] Nevertheless, consistent with concerns in the literature it suggests that researchers should be cautious in interpreting the influence of unobserved individual heterogeneity as skills (McGuinness, 2006). In this respect future research needs to explore additional omitted variables which might explain the differences between the cross-sectional and longitudinal results and would shed light on the channels through which overeducation affects labour market outcomes, key to understanding the implications for policy.

[35] Reassuringly, however, the results do not differ if instead we add the additional skill measures to capture memory and reasoning which reduce the sample considerably (results available upon request).

References

Allen, J. and Van der Velden, R. (2001). Educational mismatches versus skill mismatches: effects on wages, job satisfaction, and on-the-job search. *Oxford Economic Papers*, 53(3): 434-452.

Asadullah, M. N. and Xiao, S. (2020). The changing pattern of wage returns to education in post-reform China. *Structural Change and Economic Dynamics*, 53: 137-148.

Aslam, M., Bari, F. and Kingdon, G. (2012). Returns to schooling, ability and cognitive skills in Pakistan. *Education Economics*, 20(2): 139-173.

Battu, H. and Sloane, P. J. (2004). Over-education and ethnic minorities in Britain. *The Manchester School*, 72(4): 535-559.

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

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

Chen, Y., and Kesten, O. (2017). Chinese college admissions and school choice reforms: A theoretical analysis. *Journal of Political Economy*, 125(1): 99-139.

Chevalier, A. (2003). Measuring over-education. *Economica*, 70(279): 509-531.

Chevalier, A. and Lindley, J. (2009). Overeducation and the skills of UK graduates. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 172(2): 307-337.

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

Ferrer-i-Carbonell, A. and, Frijters, P. (2004). How important is methodology for the estimates of the determinants of happiness? *The Economic Journal*, 114(497): 641-659.

Freeman, R. B. (2010). *What does global expansion of higher education mean for the United States?*. In: American Universities in a Global Market (pp. 373-404). University of Chicago Press.

Frenette, M. (2004). The overqualified Canadian graduate: The role of the academic program in the incidence, persistence, and economic returns to overqualification. *Economics of Education Review*, 23(1): 29-45.

Green, F. and Zhu, Y. (2010). Overqualification, job dissatisfaction, and increasing dispersion in the returns to graduate education. *Oxford Economic Papers*, 62(4): 740-763.

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

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

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1): 153-161.

Heckman, J.J., Stixrud, J. and Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor economics*, 24(3): 411-482.

Kang, L., Peng, F. and Zhu, Y. (2021). Returns to higher education subjects and tiers in China: evidence from the China Family Panel Studies. *Studies in Higher Education*, 46(8): 1682-1695.

Kler, P. (2005). Graduate overeducation in Australia: A comparison of the mean and objective methods. *Education Economics*, 13(1): 47-72.

Levels, M., Van der Velden, R., and Allen, J. (2014). Educational mismatches and skills: New empirical tests of old hypotheses. *Oxford Economic Papers*, 66(4): 959-982.

Li, F., Morgan, W. J. and Ding, X. (2008). The expansion of higher education, employment and over-education in China. *International Journal of Educational Development*, 28(6): 687-697.

Li, S., Whalley, J. and Xing, C. (2014). China's higher education expansion and unemployment of college graduates. *China Economic Review*, 30: 567-582.

Liu, X. (2020). Structural changes and economic growth in China over the past 40 years of reform and opening-up. *China Political Economy*, 3(1): 19-38.

Liu, Y. (2023). Personality traits and occupational status-evidence from China. *Plos one*, 18(4), p.e0284050.

Liu, Y., Yin, L., and Guo, J. (2021). The quality of higher education and overeducation: Where should higher education funding go? *Finance Research Letters*, 41, 101824

Ma, W., Baek, J., Qi, M., Li, J. and Liu, B. (2022). The influence of overeducation on Chinese workers job satisfaction from China Household Tracking Survey (2014-2018), *International Journal of Environmental Research and Public Health*, 19, 16032.

Mateos-Romero, L. and Salinas-Jiménez, M. D. M. (2017). Skills heterogeneity among graduate workers: Real and apparent overeducation in the Spanish labor market. *Social Indicators Research,* 132(3): 1247-1264.

Mavromaras, K., McGuinness, S., O'Leary, N., Sloane, P. and Wei, Z. (2013). Job mismatches and labour market outcomes: Panel evidence on University graduates. *Economic Record*, 89: 382–395.

McGuinness, S. (2006). Overeducation in the labour market. *Journal of Economic Surveys*, 20(3): 387-418.

Mehta, A., Felipe, J., Quising, P., and Camingue, S. (2011). Overeducation in developing economies: How can we test for it, and what does it mean? *Economics of Education Review*, 30(6): 1334-1347.

Meng, X., Shen, K. and Xue, S. (2013). Economic reform, education expansion, and earnings inequality for urban males in China, 1988–2009. *Journal of Comparative Economics*, 41(1): 227-244.

Nieto, S. and Ramos, R. (2017). Overeducation, skills and wage penalty: Evidence for Spain using PIAAC data. *Social Indicators Research*, 134(1): 219-236.

Palczyńska, M. (2021). Overeducation and wages: the role of cognitive skills and personality traits. *Baltic Journal of Economics*, 21(1): 85-111.

Quinn, M. A., and 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.

Robst, J. (2008). Overeducation and college major: Expanding the definition of mismatch between schooling and jobs. *The Manchester School*, 76(4): 349–368.

Shi, Y., Yu, H., Huang, Y., Shen, X. and Guo, W. (2023) Over-education and job satisfaction among new graduates in China: A gender perspective, *Social Indicators Research*, 170: 51-74.

Sohn, K. (2010). The role of cognitive and noncognitive skills in overeducation. *Journal of Labor Research*, 31(2): 124-145.

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

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

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

Wang, C., Wan, G. and Yang, D. (2014). Income inequality in the People's Republic of China: Trends, determinants, and proposed remedies. *Journal of Economic Surveys*, 28(4): 686-708.

World Bank (2023). "School enrollment, tertiary (% gross)". The World Bank: Data. Available at: https://data.worldbank.org/indicator/SE.TER.ENRR. [Accessed 2 May 2024].

Wu, B. and Zheng, Y. (2008). *Expansion of higher education in China: Challenges and implications. The University* of Nottingham, China Policy Institute, Briefing Series, 36(1): 1-13.

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

Xing, C., Yang, P. and Li, Z. (2018). The medium-run effect of China's higher education expansion on the unemployment of college graduates. *China Economic Review*, 51: 181-193.

Yin, Y. (2004). *Disparities between urban and rural areas and among different regions in Chin*a. In: Income Disparities in China: An OECD Perspective. Paris, France: OECD Publications.

Zheng, Y., Zhang, X. and 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, 101584.

Table 1. Overeducation, and log hourly pay and job satisfaction by overeducation

	Objective	Statistical	Subjective
Overeducation (%)	36.87	46.60	46.12
Number of observations	6,170	6,170	4,508
Overeducated			
Average log hourly pay	3.099	3.222	3.055
	[0.765]	[0.824]	[0.789]
Average job satisfaction	3.544	3.559	3.531
	[0.785]	[0.761]	[0.776]
Number of observations	2,275	2,875	2,079
Not overeducated			
Average log hourly pay	3.332	3.267	3.284
	[0.769]	[0.730]	[0.760]
Average job satisfaction	3.686	3.699	3.676
	[0.734]	[0.746]	[[0.726]
Number of observations	3,895	3,295	2,429

Notes: Authors' calculation using data from the CFPS (2014-2020). See text for a description of sample construction and variable definitions. Figures in [ ] are standard deviations.

Table 2. Overeducation and log hourly pay

Panel A: Objective	(1)	(2)	(3)
Overeducation	-0.221***	-0.153***	-0.167***
	(0.023)	(0.021)	(0.021)
Personal characteristics	No	Yes	Yes
Work-related characteristics	No	No	Yes
Number of observations	6,170	6,170	6,170
Adjusted R^2	0.065	0.244	0.268
Panel B: Statistical	(1)	(2)	(3)
Overeducation	-0.003	-0.069***	-0.084***
	(0.023)	(0.021)	(0.021)
Personal characteristics	No	Yes	Yes
Work-related characteristics	No	No	Yes
Number of observations	6,170	6,170	6,170
Adjusted R^2	0.046	0.238	0.262
Panel C: Subjective	(1)	(2)	(3)
Overeducation	-0.228***	-0.254***	-0.249***
	(0.029)	(0.025)	(0.024)
Personal characteristics	No	Yes	Yes
Work-related characteristics	No	No	Yes
Number of observations	4,508	4,508	4,508
Adjusted R^2	0.061	0.250	0.273

Notes: (i) Estimates are based on a log hourly earnings equation. All models include a constant term and year fixed effects. Personal characteristics include age, age-squared, gender, ethnicity, marital status, household registration status, nature of (graduate) education and province. Work-related characteristics include (formal) contract, sector and industry (one-digit categories). (ii) Figures in ( ) are standard errors clustered at the individual level. (iii) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3. Overeducation and job satisfaction

Panel A: Objective	(1)	(2)	(3)
Overeducation	-0.136***	-0.106***	-0.078***
	(0.023)	(0.023)	(0.024)
Personal characteristics	No	Yes	Yes
Work-related characteristics	No	No	Yes
Number of observations	6,170	6,170	6,170
Adjusted R^2	0.028	0.044	0.060
Panel B: Statistical	(1)	(2)	(3)
Overeducation	-0.120***	-0.103***	-0.045*
	(0.022)	(0.023)	(0.024)
Personal characteristics	No	Yes	Yes
Work-related characteristics	No	No	Yes
Number of observations	6,170	6,170	6,170
Adjusted R^2	0.027	0.044	0.059
Panel C: Subjective	(1)	(2)	(3)
Overeducation	-0.145***	-0.147***	-0.125***
	(0.027)	(0.027)	(0.027)
Personal characteristics	No	Yes	Yes
Work-related characteristics	No	No	Yes
Number of observations	4,508	4,508	4,508
Adjusted R^2	0.023	0.044	0.064

Notes: (i) Estimates are based on a job satisfaction equation. All models include a constant term and year fixed effects. Personal characteristics include age, age-squared, gender, ethnicity, marital status, household registration status, nature of (graduate) education and province. Work-related characteristics include (formal) contract, sector and industry (one-digit categories). (ii) Figures in ( ) are standard errors clustered at the individual level. (iii) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4. Exploring the pay penalty associated with overeducation

Panel A: Objective	(1)	(2)	(3)	(4)	(5)	(6)
Overeducation	-0.170***	-0.176***	-0.170***	-0.151***	-0.143***	-0.065**
	(0.024)	(0.024)	(0.026)	(0.024)	(0.032)	(0.027)
Subject	Yes	No	No	No	Yes	No
Cognitive skills	No	Yes	No	No	Yes	No
Cognitive skill mismatch	No	No	Yes	No	Yes	No
Non-cognitive skills	No	No	No	Yes	Yes	No
Individual fixed effects	No	No	No	No	No	Yes
Number of observations	4,693	4,859	4,859	4,746	3,295	6,170
Adjusted/within R^2	0.285	0.271	0.271	0.268	0.288	0.166
Panel B: Statistical	(1)	(2)	(3)	(4)	(5)	(6)
Overeducation	-0.091***	-0.074***	-0.057**	-0.062***	-0.039	-0.008
	(0.024)	(0.024)	(0.027)	(0.024)	(0.033)	(0.027)
Subject	Yes	No	No	No	Yes	No
Cognitive skills	No	Yes	No	No	Yes	No
Cognitive skill mismatch	No	No	Yes	No	Yes	No
Non-cognitive skills	No	No	No	Yes	Yes	No
Individual fixed effects	No	No	No	No	No	Yes
Number of observations	4,693	4,859	4,859	4,746	3,295	6,170
Adjusted/within R^2	0.279	0.264	0.264	0.262	0.283	0.164
Panel C: Subjective	(1)	(2)	(3)	(4)	(5)	(6)
Overeducation	-0.274***	-0.260***	-0.257***	-0.227***	-0.263***	-0.054
	(0.027)	(0.027)	(0.022)	(0.028)	(0.033)	(0.061)
Subject	Yes	No	No	No	Yes	No
Cognitive skills	No	Yes	No	No	Yes	No
Cognitive skill mismatch	No	No	Yes	No	Yes	No
Non-cognitive skills	No	No	No	Yes	Yes	No
Individual fixed effects	No	No	No	No	No	Yes

Number of observations	3,647	3,943	3,943	3,568	2,724	4,508
Adjusted/within R^2	0.289	0.273	0.274	0.267	0.291	0.166

Notes: (i) Estimates are based on a log hourly pay equation. All models include a constant term, year fixed effects and personal and work-related characteristics, although time-invariant characteristics are excluded from column (6). Personal characteristics include age, age-squared, gender, ethnicity, marital status, household registration status, nature of (graduate) education and province. Work-related characteristics include (formal) contract, sector and industry (one-digit categories). Controls for subject include Economics, Law, Education, Literature, Science, Engineering, Agronomy, Medicine, Management and Other. Controls for cognitive skills include measures of literacy and numeracy. Controls for cognitive skill mismatch include measures relating to literacy and numeracy skill mismatch. Non-cognitive skills include locus of control, conscientiousness, extroversion, agreeableness, openness and neuroticism. (ii) Figures in ( ) are standard errors clustered at the individual level. (iii) Reported are the adjusted R^2 in columns (1)-(5) and within R^2 in column (6). (iv) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5. Exploring the job satisfaction penalty associated with overeducation

Panel A: Objective	(1)	(2)	(3)	(4)	(5)	(6)
Overeducation	-0.105***	-0.098***	-0.075***	-0.074***	-0.049	-0.004
	(0.028)	(0.028)	(0.029)	(0.028)	(0.036)	(0.034)
Subject	Yes	No	No	No	Yes	No
Cognitive skills	No	Yes	No	No	Yes	No
Cognitive skill mismatch	No	No	Yes	No	Yes	No
Non-cognitive skills	No	No	No	Yes	Yes	No
Individual fixed effects	No	No	No	No	No	Yes
Number of observations	4,693	4,859	4,859	4,746	3,295	6,170
Adjusted/within R^2	0.070	0.063	0.064	0.105	0.121	0.007
Panel B: Statistical	(1)	(2)	(3)	(4)	(5)	(6)
Overeducation	-0.038	-0.068***	-0.040	-0.056**	-0.010	0.007
	(0.028)	(0.027)	(0.026)	(0.027)	(0.036)	(0.033)
Subject	Yes	No	No	No	Yes	No
Cognitive skills	No	Yes	No	No	Yes	No
Cognitive skill mismatch	No	No	Yes	No	Yes	No
Non-cognitive skills	No	No	No	Yes	Yes	No
Individual fixed effects	No	No	No	No	No	Yes
Number of observations	4,693	4,859	4,859	4,746	3,295	6,170
Adjusted/within R^2	0.067	0.062	0.063	0.104	0.120	0.069
Panel C: Subjective	(1)	(2)	(3)	(4)	(5)	(6)
Overeducation	-0.156***	-0.146***	-0.139***	-0.102***	-0.139***	-0.176**
	(0.032)	(0.030)	(0.030)	(0.031)	(0.037)	(0.080)
Subject	Yes	No	No	No	Yes	No
Cognitive skills	No	Yes	No	No	Yes	No
Cognitive skill mismatch	No	No	Yes	No	Yes	No
Non-cognitive skills	No	No	No	Yes	Yes	No
Individual fixed effects	No	No	No	No	No	Yes

Number of observations	3,647	3,943	3,943	3,568	2,724	4,508
Adjusted/within R^2	0.074	0.065	0.067	0.101	0.126	0.064

Notes: (i) Estimates are based on a job satisfaction equation. All models include a constant term, year fixed effects and personal and work-related characteristics, although time-invariant characteristics are excluded from column (6). Personal characteristics include age, age-squared, gender, ethnicity, marital status, household registration status, nature of (graduate) education and province. Work-related characteristics include (formal) contract, sector and industry (one-digit categories). Controls for subject include Economics, Law, Education, Literature, Science, Engineering, Agronomy, Medicine, Management and Other. Controls for cognitive skills include measures of literacy and numeracy. Controls for cognitive skill mismatch include measures relating to literacy and numeracy skill mismatch. Non-cognitive skills include locus of control, conscientiousness, extroversion, agreeableness, openness and neuroticism. (ii) Figures in ( ) are standard errors clustered at the individual level. (iii) Reported are the adjusted R^2 in columns (1)-(5) and within R^2 in column (6). (iv) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Online Appendix: Overeducation, Earnings and Job Satisfaction among Graduates in China

Table A1. Number of waves per individual

Total number of waves present	Number of individuals	% of observations
1	1,438	44.92
2	869	27.14
3	582	18.18
4	312	9.76

Notes: Authors' calculation using data from the CFPS (2014-2020). See text for a description of sample construction.

Table A2. Overlap between overeducation measures

	Objective		Statistical		Subjective	
	Overeducated	Non-overeducated	Overeducated	Non-overeducated	Overeducated	Non-overeducated
Objective						
Overeducated	-	-	1,758	517	988	745
[% of objective overeducated]			[77.27]	[22.73]	[57.01]	[42.99]
Non-overeducated	-	-	1,117	2,778	1,091	1,684
[% of objective non-overeducated]			[28.68]	[71.32]	[39.32]	[60.68]
Statistical						
Overeducated	1,758	1,117	-	-	1,217	947
[% of statistical overeducated]	[61.15]	[38.85]			[56.24]	[43.76]
Non-overeducated	517	2,778	-	-	862	1,482
[% of statistical non-overeducated]	[15.69]	[84.31]			[36.77]	[63.23]
Subjective						
Overeducated	988	1,091	1,217	862	-	-
[% of subjective overeducated]	[47.52]	[52.48]	[58.54]	[41.46]		
Non-overeducated	745	1,684	947	1,482	-	-
[% of subjective non-overeducated]	[30.67]	[69.33]	[38.99]	[61.01]		

Notes: Authors' calculation using data from the CFPS (2014-2020). See text for a description of sample construction and variable definitions.

Table A3. Transition rates in overeducation

Panel A. Objective	Overeducated	Non-overeducated	Total
Overeducated	728	344	1,072
	[67.91%]	[32.09%]	[100%]
Non-overeducated	280	1,617	1,897
	[14.76%]	[85.24%]	[100%]
Total	1,008	1,961	2,969
	[33.95%]	[66.05%]	[100%]
Panel B. Statistical	Overeducated	Non-overeducated	Total
Overeducated	981	450	1,431
	[68.55%]	[31.45%]	[100%]
Non-overeducated	230	1,308	1,538
	[14.95%]	[85.05%]	[100%]
Total	1,211	1,758	2,969
	[40.79%]	[59.21%]	[100%]
Panel C. Subjective	Overeducated	Non-overeducated	Total
Overeducated	945	85	1,030
	[91.75%]	[8.25%]	[100%]
Non-overeducated	59	1,046	1,105
	[5.34%]	[94.66%]	[100%]
Total	1,004	1,131	2,135
	[47.03%]	[52.97%]	[100%]

Notes: Authors' calculation using data from the CFPS (2014-2020). See text for a description of sample construction and variable definitions. Transitions are defined as changes in overeducation status within individuals observed within the panel.

Table A4. Definitions of non-cognitive skills

Item	Related questions
Locus of control	Wealth reflects personal achievement
	Hard work pays off
	Intelligence pays off
	The social relationship is more important than hard work
	There are great opportunities for me to improve my living standards
	I am confident in the future
Openness	Having originality and creativity
	Pay attention to the experience of art and aesthetics
	Be imaginative
Conscientiousness	Be rigorous and serious
	Often be very lazy
	Do jobs efficiently
Extroversion	Love to talk
	Be cheerful and sociable
	Be reserved and conservative
Agreeableness	Be tolerant of nature
	Sometimes be rude to others
	Be considerate of others and kind to almost everyone
Neuroticism	Often be worried
	Easy to be nervous
	Often be relaxed and able to deal with pressure

Source: CFPS.

Table A5. Means of explanatory variables by overeducation

	Objective		Statistical		Subjective	
	Overeducated	Non-overeducated	Overeducated	Non-overeducated	Overeducated	Non-overeducated
Personal characteristics						
Age	33.485	34.754	33.536	34.941	33.879	34.315
Male	0.594	0.497	0.613	0.463	0.550	0.512
Minority ethnicity	0.049	0.059	0.049	0.060	0.054	0.061
Married	0.734	0.787	0.741	0.791	0.748	0.740
Non-agricultural hukou	0.654	0.714	0.681	0.701	0.646	0.711
University	0.339	0.553	0.405	0.534	0.474	0.437
Postgraduate	0.022	0.066	0.105	0.002	0.074	0.027
Work-related characteristics						
Formal contract	0.747	0.805	0.776	0.790	0.726	0.807
Public sector	0.425	0.572	0.443	0.583	0.457	0.521
Subject						
Economics	0.162	0.135	0.144	0.145	0.149	0.143
Law	0.067	0.043	0.058	0.046	0.059	0.048
Education	0.032	0.118	0.023	0.141	0.057	0.098
Literature	0.061	0.105	0.059	0.115	0.100	0.078
Science	0.067	0.071	0.074	0.065	0.066	0.073
Engineering	0.248	0.196	0.260	0.177	0.225	0.202
Agronomy	0.026	0.018	0.027	0.015	0.025	0.019
Medicine	0.026	0.090	0.076	0.057	0.045	0.073
Management	0.186	0.130	0.165	0.138	0.156	0.152
Others (including Arts, Philosophy and History)	0.126	0.096	0.115	0.101	0.117	0.114
Cognitive skills						
Literacy	28.645	29.749	29.054	29.593	29.060	29.407
Numeracy	15.696	16.972	16.271	16.703	16.355	16.604

Cognitive skill mismatch							
	Overskill (literacy)	0.252	0.038	0.241	0.010	0.155	0.095
	Overskill (numeracy)	0.445	0.254	0.445	0.220	0.368	0.301
Non-cognitive skills							
	Locus of control	3.528	3.546	3.555	3.526	3.512	3.552
	Conscientiousness	3.708	3.815	3.747	3.801	3.751	3.795
	Extroversion	3.206	3.219	3.214	3.214	3.164	3.248
	Agreeableness	3.451	3.448	3.461	3.439	3.448	3.448
	Openness	3.246	3.335	3.291	3.313	3.203	3.362
	Neuroticism	3.163	3.173	3.187	3.155	3.158	3.190
Year							
	2014	0.224	0.185	0.250	0.155	0.264	0.268
	2016	0.263	0.267	0.278	0.254	0.258	0.258
	2018	0.255	0.266	0.239	0.282	0.246	0.231
	2020	0.258	0.282	0.232	0.309	0.231	0.243

Notes: Authors' calculation using data from the CFPS (2014-2020). Figures relate to relevant estimation sample. See text for a description of sample construction. Descriptive statistics for province and industry are available upon request.

Table A6. Full set of coefficient estimates, hourly pay and job satisfaction

	Hourly pay			Job satisfaction		
	Objective	Statistical	Subjective	Objective	Statistical	Subjective
	(1)	(2)	(3)	(4)	(5)	(6)
Overeducated	-0.167***	-0.084***	-0.249***	-0.078***	-0.045**	-0.125***
	(0.021)	(0.021)	(0.024)	(0.024)	(0.024)	(0.027)
Age	-0.001	-0.000	0.024	-0.040***	-0.039***	-0.039**
	(0.014)	(0.014)	(0.017)	(0.014)	(0.014)	(0.017)
Age squared	0.000	0.000	-0.000	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Male	0.181***	0.186***	0.182***	-0.004	-0.001	0.005
	(0.021)	(0.021)	(0.025)	(0.024)	(0.024)	(0.027)
Minority ethnicity	0.000	0.002	-0.041	0.094*	0.095*	0.068
	(0.052)	(0.052)	(0.061)	(0.054)	(0.054)	(0.062)
Married	0.011	0.016	0.006	0.084***	0.087***	0.079**
	(0.027)	(0.027)	(0.032)	(0.029)	(0.029)	(0.033)
Non-agricultural hukou	0.049*	0.045*	-0.014	-0.015	-0.016	-0.072**
	(0.026)	(0.026)	(0.030)	(0.027)	(0.027)	(0.031)
University	0.230***	0.259***	0.290***	0.024	0.038	0.026
	(0.022)	(0.022)	(0.026)	(0.025)	(0.025)	(0.029)
Postgraduate	0.477***	0.570***	0.610***	-0.002	0.046	0.093
	(0.052)	(0.054)	(0.061)	(0.049)	(0.051)	(0.057)
Formal contract	0.159***	0.165***	0.148***	0.046	0.049*	0.063**
	(0.026)	(0.026)	(0.029)	(0.028)	(0.028)	(0.031)
Public sector	-0.018	-0.025	-0.063**	-0.016	-0.020	-0.037
	(0.025)	(0.026)	(0.029)	(0.029)	(0.029)	(0.033)
Constant	2.224***	2.144***	2.105***	4.042***	4.008***	3.576***
	(0.365)	(0.362)	(0.378)	(0.371)	(0.369)	(0.733)
Number of observations	6170	6170	4508	6170	6170	4508
Adjusted R^2	0.268	0.262	0.273	0.060	0.059	0.064

Notes: (i) Estimates are based on log hourly pay (columns (1)-(3)) and job satisfaction (columns (4)-(6)) equations. Controls for year, province and industry are included but coefficients are not reported. (ii) Figures in ( ) are standard errors clustered at the individual level. (iii) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A7. Overeducation and facets of job satisfaction

Panel A. Pay	Objective	Statistical	Subjective
Overeducated	-0.065**	-0.047	-0.162***
	(0.032)	(0.031)	(0.038)
Number of observations	4,941	4,941	3,308
Adjusted R^2	0.068	0.068	0.070
Panel B. Safety	Objective	Statistical	Subjective
Overeducated	-0.125***	-0.184***	-0.127***
	(0.030)	(0.030)	(0.036)
Number of observations	4,940	4,940	3,308
Adjusted R^2	0.101	0.106	0.099
Panel C. Environment	Objective	Statistical	Subjective
Overeducated	-0.117***	-0.115***	-0.169***
	(0.031)	(0.030)	(0.036)
Number of observations	4,941	4,941	3,308
Adjusted R^2	0.091	0.091	0.091
Panel D. Working hours	Objective	Statistical	Subjective
Overeducated	-0.028	-0.103***	-0.040
	(0.034)	(0.034)	(0.040)
Number of observations	4,941	4,941	3,308
Adjusted R^2	0.050	0.052	0.050
Panel E. Promotion	Objective	Statistical	Subjective
Overeducated	-0.114***	-0.080**	-0.171***
	(0.036)	(0.035)	(0.043)
Number of observations	4,484	4,484	2,987
Adjusted R^2	0.029	0.027	0.037

Notes: (i) Estimates are based on an equation for each facet of job satisfaction. Data relate to 2016-2020. All models include a constant term, year fixed effects and personal and work-related characteristics. Personal characteristics include age, age-squared, gender, ethnicity, marital status, household registration status, nature of (graduate) education and province. Work-related characteristics include (formal) contract, sector and industry (one-digit categories). (ii) Figures in ( ) are standard errors clustered at the individual level. (iii) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A8. Overeducation and hourly pay, sensitivity analysis

	Dropping oversampled observations	Controlling for occupation	Controlling for firm size	Controlling for job tenure	Controlling for institutional quality	Distinguishing matched and undereducated	Only full-time employees	Heckman selection correction	Excluding pay outliers
Panel A: Objective	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Overeducation	-0.163***	-0.095***	-0.177***	-0.163***	-0.187***	-	-0.158***	-0.166***	-0.165***
	(0.027)	(0.030)	(0.021)	(0.022)	(0.025)		(0.020)	(0.020)	(0.020)
Number of observations	3,966	6,170	5,232	5,527	4,187	-	5,984	6,046	6,152
Adjusted R^2	0.237	0.270	0.306	0.268	0.269	-	0.316	-	0.291
Panel B: Statistical	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Overeducation	-0.090***	-0.039	-0.091***	-0.082***	-0.122***	-0.042*	-0.082***	-0.085***	-0.083***
	(0.026)	(0.029)	(0.021)	(0.022)	(0.026)	(0.022)	(0.020)	(0.020)	(0.020)
Number of observations	3,966	6,170	5,232	5,527	4,187	6,170	5,984	6,046	6,152
Adjusted R^2	0.232	0.269	0.299	0.262	0.263	0.265	0.310	-	0.285
Panel C: Subjective	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Overeducation	-0.257***	-0.230***	-0.233***	-0.256***	-0.241***	-0.228***	-0.247***	-0.248***	-0.240***
	(0.031)	(0.025)	(0.024)	(0.025)	(0.030)	(0.025)	(0.023)	(0.021)	(0.023)
Number of observations	2,915	4,508	3,716	4,244	3,164	4,508	4,360	4,432	4,492
Adjusted R^2	0.256	0.278	0.310	0.277	0.272	0.278	0.319	-	0.301

Notes: (i) Estimates are based on a log hourly pay equation. All models include a constant term, year fixed effects and personal and work-related characteristics. Personal characteristics include age, age-squared, gender, ethnicity, marital status, household registration status, nature of (graduate) education and province. Work-related characteristics include (formal) contract, sector and industry (one-digit categories). Column (1) removes observations that result from the oversampling of selected provinces in the survey design. Column (2) additionally controls for occupation (one-digit categories). Column (3) additionally controls for the log of the number of employees in the firm and column (4) similarly controls for job tenure (and tenure squared). For undergraduates, column (5) controls for institution quality (defined as Key, Ordinary Key, Ordinary and Other) and column (6) controls for undereducation. Column (7) restricts the sample to those who work a minimum of 24 hours per week. Estimates from a Heckman selection model which accounts for non-random selection into paid employment are provided in column (8). Caring responsibilities are used as exclusion restrictions, proxied by the number of children aged under 14 years old and the number of people aged more than 65 years old. Column (9) excludes pay outliers defined as more than 10 times the top percentile and half the bottom percentile within each year. (ii) Figures in ( ) are standard errors clustered at the individual level. (iii) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A9. Overeducation and job satisfaction, sensitivity analysis

	Dropping oversampled observations	Controlling for occupation	Controlling for firm size	Controlling for job tenure	Controlling for institutional quality	Distinguishing matched and undereducated	Only full-time employees	Heckman selection correction	Controlling for hourly pay
Panel A: Objective	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Overeducation	-0.096***	-0.081**	-0.076***	-0.054**	-0.099***	-	-0.072***	-0.078***	-0.050**
	(0.031)	(0.037)	(0.026)	(0.025)	(0.030)		(0.025)	(0.022)	(0.024)
Number of observations	3,966	6,170	5,232	5,527	4,187	-	5,984	6,046	6,170
Adjusted R^2	0.063	0.065	0.052	0.058	0.068	-	0.059	-	0.081
Panel B: Statistical	(1)	(2)	(3)	(4)	(5)	(7)	(8)	(9)	(6)
Overeducation	-0.051*	-0.038	-0.049*	-0.035	-0.066**	-0.028	-0.044*	-0.050**	-0.031
	(0.031)	(0.033)	(0.026)	(0.025)	(0.030)	(0.026)	(0.024)	(0.022)	(0.023)
Number of observations	3,966	6,170	5,232	5,527	4,187	6,170	5,984	6,046	6,170
Adjusted R^2	0.061	0.064	0.051	0.058	0.066	0.059	0.058	-	0.081
Panel C: Subjective	(1)	(2)	(3)	(4)	(5)	(7)	(8)	(9)	(6)
Overeducation	-0.110***	-0.111***	-0.146***	-0.112***	-0.137***	-0.120***	-0.129***	-0.122***	-0.087***
	(0.034)	(0.028)	(0.029)	(0.028)	(0.033)	(0.028)	(0.028)	(0.023)	(0.027)
Number of observations	2,915	4,508	3,716	4,244	3,164	4,508	4,360	4,432	4,508
Adjusted R^2	0.063	0.069	0.055	0.063	0.077	0.064	0.064	-	0.082

Notes: (i) Estimates are based on a log hourly pay equation. All models include a constant term, year fixed effects and personal and work-related characteristics. Personal characteristics include age, age-squared, gender, ethnicity, marital status, household registration status, nature of (graduate) education and province. Work-related characteristics include (formal) contract, sector and industry (one-digit categories). Column (1) removes observations that result from the oversampling of selected provinces in the survey design. Column (2) additionally controls for occupation (one-digit categories). Column (3) additionally controls for the log of the number of employees in the firm and column (4) similarly includes job tenure (and tenure squared). For undergraduates, column (5) controls for institution quality (defined as Key, Ordinary Key, Ordinary and Other). Column (6) additionally controls for undereducation. Column (7) restricts the sample to those who work a minimum of 24 hours per week. Estimates from a Heckman selection model which accounts for non-random selection into paid employment are provided in column (8). Caring responsibilities are used as exclusion restrictions, proxied by the number of children aged under 14 years old and the number of people aged more than 65 years old. Column (9) additionally controls for log hourly pay. (ii) Figures in ( ) are standard errors clustered at the individual level. (iii) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.