Does Education Level Affect Hourly Pay in the UK?

1. Intro

Income inequality is a crucial issue in the UK and an issue that it has faced since it began, increasing rapidly following the Industrial Revolution. A person's income plays a crucial role in determining their quality of life and raising their standard of living.

Educational qualifications are widely recognised as being a key determinant of earnings, leading to assumptions that there are wage gaps based on education level. This project's analysis uses data specific to the UK from the Understanding Society Survey and the purpose of this project is to explore the ways in which educational attainment is related to hourly pay.

Hypotheses:

H_0: Education level does not have an effect on hourly pay

H_1: Education level has an effect on hourly pay

2. Background

The correlation between educational attainment and pay has been well documented through various scholarly journals. The Human Capital Theory forms the basis for this, suggesting more educational investment increases productivity and labour value in the market. According to this, higher education leads to higher wages as qualifications signal skill to employers. (Becker, 1964)

Longhi and Platt (2008) add another layer by mentioning education as one of the most significant aspects of pay, identifying individuals with higher qualifications earn significantly more even after controlling for gender, age and occupation, showing the importance of controlling for these aspects in identifying wage gaps.

Hills (2010) provides a more comprehensive study of income inequality within the UK and provides a large statistical backdrop for investigating wage differences. The report shows that although other factors such as religion, ethnicity and disability impact pay, education is an overreaching factor, and it is noted that differences in qualifications contribute to unequal pay, and it remains one of the main symbols of inequality in UK markets.

Additionally, Blundell et al. (2005) used UK panel data to estimate returns on investment to levels of education, identifying that completing A-levels returned 24%, university degrees lead to earnings premiums of 27-48%. Furthermore, Britton et al (2022) contributes to this by showing further differences between degrees, with those with a first-class earning 13% more than those with a 2:1.

Crawford et al (2019) Explores GCSE qualifications, noting that small increases in results lead to large increases in lifetime earnings. This supports the idea that wage gaps start early from leaving school and have significant impacts later on in the labour market, supporting the idea that education is linked to pay and lifetime earnings.

3. Methods

The dataset used in this analysis is from the Understanding Society survey, with a sample consisting of over 17,000 working age respondents (N=17073). In this analysis we are using highest educational qualification as our independent variable, and pay as our dependant variable. Highest educational qualification is nominal, while pay is continuous. First, we created an hourly pay variable to control for differences in hours worked (h_jbhrpay). To obtain this variable, we divided h_paygu_dv by 4.33 (amount of weeks in a month) and then further by the h_jbhrs variable to obtain hourly pay. The formula for this computation was (h_paygu_dv/4.33) /h_jbhrs to combine the steps of calculating weekly pay and then hourly pay into one formula.

In order to improve the clarity of results and simplify the analysis, the default education variable (h_hiqual_dv) was simplified using a transformation into (h_hiqual_dv_grouped) from six categories into three, splitting different levels of educational achievement into: ‘Higher Education’ = Degrees (1 and 2), ‘School Education’ = A-level/GCSE (3 and 4), ‘No Education’ = Alternative/No qualifications(and 6). This transformation allows for increased readability and clarity in results for interpreting the wage gap between those with higher-education and those with school/no education.

The next step to identify correlations between education and pay is to create a set of descriptive statistics using frequency tables and a box plot. This helps us to identify anomalies in the statistics by selecting ‘label outliers’ and to identify the means and medians. From the frequency table, the minimum hourly pay variables appeared impossibly low. The box plot further revealed other extreme outliers of thousands of pounds per hour. In response to the anomalous data, (h_clean_jbhrpay) was generated by removing values below £3 per hour and above £300 using a transform. By cleaning this data, it ensures that the regression result will not be affected by extreme outliers or incorrect values. This cleaned variable was used in all subsequent analysis.

To support the main regression analysis, a binary value was created from h_clean_jbhrpay using a transformation to create (lowwage/highwage), where any pay <£7.55 is considered ‘lowwage’ (below 2025 UK minimum wage) and anything at or above is considered highwage. This is so that we can use contingency tables and chi-squared tests to support the main regression and identify association between education and likelihood of low pay.

A histogram of h_clean_jbhrpay revealed a highly right-skewed distribution despite earlier data cleaning. To address this and improve model fit, a log transformed version was created (ln_hourlypay). This transformation helps to satisfy the linear regression assumption that the data is normally distributed.

ln_hourlypay was then run in a log-linear regression for comparison to standard h_clean_jbhrpay to assess for differences in normality of residuals, collinearity and homoscedasticity. Ln_hourlypay was chosen due to improved model fit, explanatory power and correct for the skewness in hourly pay and improve interpretability of results. The adjusted R^2 increased and the residuals more closely aligned with normal distribution through a Q-Q plot.

Residual plots were also used to assess for homoscedasticity and model fit. The residuals vs fitted values plot showed no pattern or funnelling, suggesting that the constant variance assumption is met. The residuals vs predicted plot indicated a linear relationship, further supporting the suitability of the model. These diagnostics confirm that all assumptions were met alongside the improved adjusted R^2 and improved residual normality from the Q-Q plot.

A multivariate regression model was then set up using the log-transformed variables, with ln_hourlypay as dependent variable and h_hiqual_dv_grouped as independent variable, furthermore sex (h_sex) and manager duties (h_jbmngr) were introduced as controls.

4. Results

This section presents the results of the regression examining the relationship between education level and pay while controlling for gender and managerial status.

Table 1 displays that N=17073 people in the sample, and out of that 6118 fell under ‘Higher Education’, 9933 under ‘School Education’ and 1022 under ‘No Education’, making A-level/GCSE qualifications the most prominent. Table 2 shows us that the mean (£14.60) hourly pay is moderately higher than the median (£11.80), this shows us that a few higher earners are driving the mean up whilst the majority fall below that mean. This is shown with outliers labelled (Figure 1). Additionally, we can see that those who have higher education have a significantly higher mean (£19.10) than those with school education (£12.40) compared to a comparatively smaller difference between school and no education (£11.10, £1.30 difference).

To explore if there is a link between education and likelihood of being low paid, a contingency table was created with the binary variable we made earlier (lowwage/highwage) and our grouped education variable. As can be seen from the results in tables 3 and 4, a chi-square(X^2) test found there was significant association between education and pay (X^2 = 748, p <.001). This high association and low p-value reflect the strong association between the two variables.

The quantity of individuals under the ‘lowwage’ category was lowest for the higher education group (6.0%) compared to 21.8% for those under school education and 24.8% for those with no qualification. The same trend as before appears, with strong pay disparities appearing between higher education and other groups, but smaller disparities between the other groups. This is further reinforced by our earlier descriptives. These findings suggest that individuals without formal qualifications are more than four times more likely to be low-paid in comparison to those with higher education, which is further illustrated by the bar plot (Figure 2) where the low-wage quantity decreases as education rises.

A log-linear regression was also done to examine the effect of education on hourly pay whilst implementing controls for gender and managerial responsibilities. This is to account for differences in pay by sex and the fact that those with managerial duties are paid more. The model was statistically significant as seen in table 5 with (F-test(n=17061) =1318), $p < .001$ and an adjusted 0.278 R^2, indicating approximately 27.8% of variance in ln_hourpay is explained by this model.

Based on the results of this log-linear regression, we reject the null hypothesis (H_0) that education does not have an effect on hourly pay. As seen by tables 5 and 6, education was a significant predictor of pay. Those under 'school-level' education made roughly 28.6% less per hour than those under 'higher education', and those under 'no education' earned 33.5% less. From the controls, we can see that females earned 12.2% less than males and those in non-managerial positions made 20.5% less than managers. Ultimately, these reports strongly reinforce the idea that higher education is strongly associated with findings even with controls.

Before accepting the regression results, it was checked that the assumptions of linear regression were met. The Q-Q plot (Figure 3) shows that the residuals follow a normal distribution with a few deviations at the extreme ends. The plot of residuals vs fitted values shows no funnel shape, which suggests the variance is constant (homoscedasticity) (Figures 4 and 5). The VIF for all predictors was close to 1 (Table 7), ruling out multicollinearity. This suggests that the assumptions were met well enough to trust the model.

The findings of this regression are overall consistent with the earlier descriptive statistics that we collected and the conducted contingency analysis (those with higher education had significantly lower rates of low-wage employment and significantly higher average job pay). The regression reinforces this pattern but in a more precise way as it controlled for other factors like sex and managerial duties. That gap in overall hourly pay between those in the 'higher education' group and those in the 'school education' and 'no education' group were substantial, with those receiving only school education earning roughly 28.6% less per hour, and those without qualifications earning around 33.5% less (Table 6). Our chi-square results using the lowwage/highwage variable align with this, showing that education reliably predicts both average pay and the likelihood of being low paid. This was mirrored in the contingency analysis (Table 3) where only 6% of individuals with higher education fell under low wage compared to over 20% for those without it.

Although this model shows very strong associations and is enough to reject the null hypothesis, it is based on purely observational data. Other factors such as job, industry and experience were not included.

5. Conclusion

This analysis answers the question 'Does education affect the level of hourly pay in the UK?'. From our analysis we can verify that education has a large and consistent effect on pay even after controlling for managerial roles and sex, leading to us rejecting the null hypothesis. This was further reinforced using the binary lowwage variable alongside the log-transformed and cleaned hourly pay variable with contingency tables, linear statistics and a log-linear multivariate regression.

These findings highlight the importance of education in UK pay outcomes, especially in avoiding low-wage employment. Although factors like job type and experience were not included, the evidence suggests that education plays a key role in creating economic opportunity.

Word Count: 1903

6. References

Becker, G.S. (1964) *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. Chicago: University of Chicago Press

Blundell, R., Dearden, L. and Sianesi, B. (2005) 'Evaluating the impact of education on earnings in the UK: models, methods and results from the NCDS', *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 168(3)

Britton, J., Dearden, L. and Waltmann, B. (2022) *How much does it pay to get good grades at university?*. Department for Education

Crawford, C., Macmillan, L. and Vignoles, A. (2019) *GCSE attainment and lifetime earnings*. Department for Education.

Hills, J. (2010) *An anatomy of economic inequality in the UK: Report of the National Equality Panel*. London: London School of Economics, CASE report 60

Longhi, S. and Platt, L. (2008) *Pay gaps across equalities areas: An analysis of pay gaps and pay penalties by sex, ethnicity, religion, disability, sexual orientation and age using the Labour Force Survey*. Institute for Social and Economic Research, University of Essex

7. Tables

Table 1 Descriptives of cleaned hourly pay split by grouped educational qualifications

Descriptives

	h_hiqual_dv_grouped	h_clean_jbhrpay
N	**Higher Education**	6118

Descriptives

	h_hiqual_dv_grouped	h_clean_jbhrpay
	School Education	9933
	No Education	1022
Mean	**Higher Education**	19.1
	School Education	12.4
	No Education	11.1
Median	**Higher Education**	16.6
	School Education	10.4
	No Education	9.59

Table 2 Descriptives of cleaned hourly pay

Descriptives

	h_clean_jbhrpay
N	17858
Mean	14.6
Median	11.8

Table 3 Contingency table showing links between grouped educational qualifications and the binary lowwage/highwage variable

Contingency Tables

		lowwage/highwage		
h_hiqual_dv_grouped		**lowwage**	**highwage**	**Total**
Higher Education	Observed	368	5750	6118
	Expected	999	5119	6118
	% within row	6.0%	94.0%	100.0%
School Education	Observed	2166	7767	9933
	Expected	1621	8312	9933
	% within row	21.8%	78.2%	100.0%
No Education	Observed	253	769	1022
	Expected	167	855	1022
	% within row	24.8%	75.2%	100.0%
Total	Observed	2787	14286	17073
	Expected	2787	14286	17073
	% within row	16.3%	83.7%	100.0%

Table 4 Chi-squared (X^2) tests of grouped education and the lowwage/highwage variable

χ^2 Tests

	Value	df	p
χ^2	748	2	<.001
N	17073		

Table 5 Log-linear regression model fit output including adjusted R^2 values and model tests

Model Fit Measures

			Overall Model Test			
Model	**R^2**	**Adjusted R^2**	**F**	**df1**	**df2**	**p**
1	0.279	0.278	1318	5	17061	<.001

Note. Models estimated using sample size of N=17067

Table 6 Model coefficients of log-linear regression output including controls for managerial status and sex. (higher education set as reference level)

Model Coefficients - ln_hourpay

			95% Confidence Interval			
Predictor	**Estimate**	**SE**	**Lower**	**Upper**	**t**	**p**
Intercept[a]	3.134	0.00867	3.117	3.151	361.7	<.001
h_hiqual_dv_grouped:						
School Education – Higher Education	-0.337	0.00754	-0.352	-0.322	-44.7	<.001
No Education – Higher Education	-0.408	0.01548	-0.438	-0.377	-26.3	<.001
h_sex:						
Female – Male	-0.130	0.00703	-0.144	-0.116	-18.5	<.001
h_jbmngr:						
Foreman/supervisor – Manager	-0.229	0.01186	-0.253	-0.206	-19.3	<.001
NOT manager or supervisor – Manager	-0.434	0.00874	-0.451	-0.417	-49.7	<.001

[a] Represents reference level

Table 7 Collinearity statistics of log linear regression model (assumption checks)

Collinearity Statistics

	VIF	Tolerance
h_hiqual_dv_grouped	1.01	0.987
h_sex	1.01	0.993
h_jbmngr	1.02	0.985

Figure 1 Boxplot of cleaned hourly pay (h_clean_jbhrpay) with outliers labelled

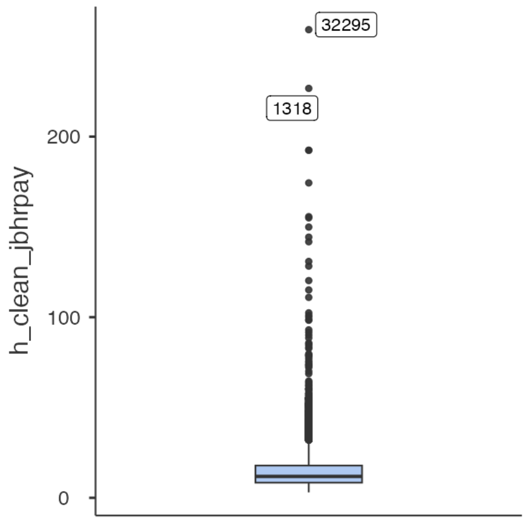

Figure 2 Bar plot of lowwage/highwage binary variable split by education

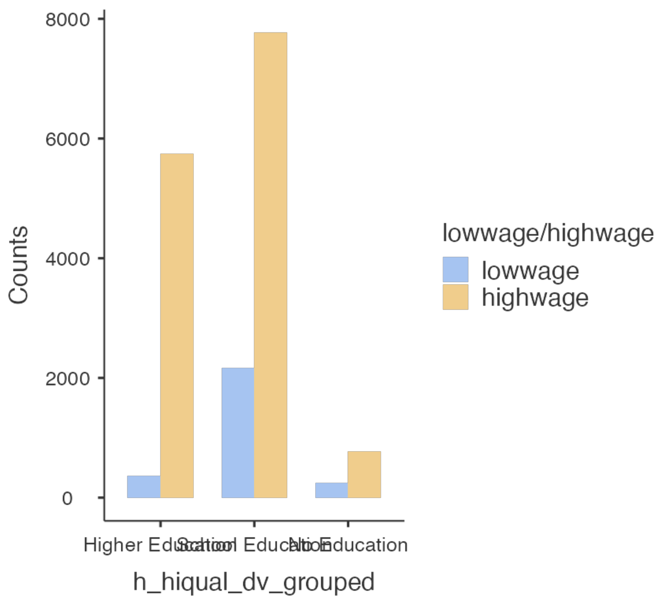

Figure 3 Q-Q plot of log-linear regression residuals (assumption check)

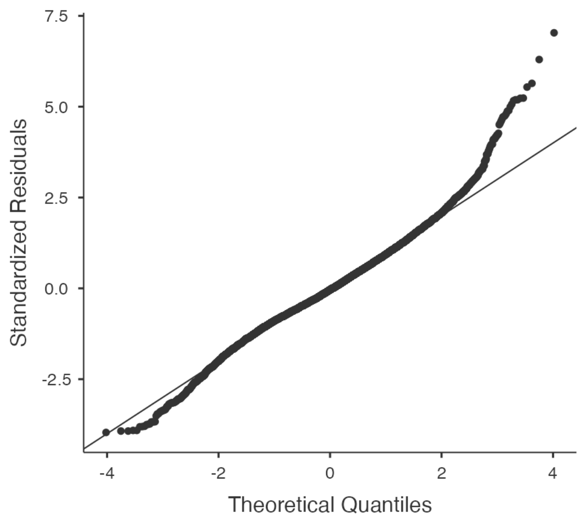

Figure 4 Residuals vs fitted results of log-linear regression (assumption check)

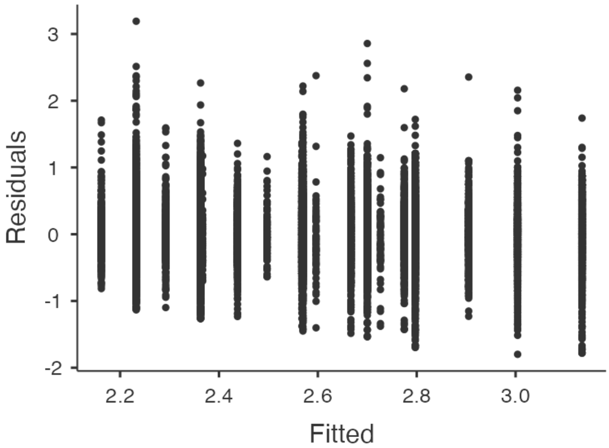

Figure 5 Predicted residuals vs ln_hourpay of log-linear regression (assumption check)