

From the Dinner Table to the Labour Market: Family Cohesion during Youth and Future Adult Wages.

Muaadh Tanveer

Professional Economist BSc and Apprenticeship Level 6

School of Economics

University of Kent, 2026

Abstract

Drawing on human and social capital theories where family relations activate cultural and intellectual endowments, this paper investigates the long-run impact of adolescent family cohesion on future adult earnings using Next Steps cohort data tracking 15,770 individuals in England from age 14 to 32. Constructing a continuous Family Cohesion Factor (FCF) through Principal Axis Factoring across youth-reported dimensions, the empirical strategy employs baselines OLS log-wage regressions alongside WLS, Heckman Two-Step corrections, and School Fixed-Effects models to adjust for non-random attrition, sample selection, and time-invariant unobservable characteristics. Across all specifications, a one standard deviation increase in FCF at age 14 is associated with a statistically significant 2.2-3.3% increase in gross weekly wages at age 32. Oster bound testing confirms robustness against omitted variable bias, while mediation analysis demonstrates that the association operates primarily through the cohort member's attitude towards school at age 16.

AI Statement

I acknowledge the use of generative AI in drafting and code development in this paper. However, the work reported remains my own.

Acknowledgements

All praise is due to Allah, by Whose blessings all good actions are brought to completion. I express my appreciation to all those who assisted in this endeavour, whether through practical support or inspiration. I am conscious of standing on the shoulders of prior thinkers, inheriting insights across centuries, continents, and disciplines – from those who when faced with obscurity, retreated in quiet to supplicate for intellectual opening, to those who called all corners of society to their shared duty of cultivation. I ask the Most High to render this contemporary effort a source of enduring benefit, and a seed for a new generation of economic thinkers building upon these grassroots.

1. Introduction

For most children in England, the family home is the first and most formative institution they will encounter. The Social Mobility Commission's State of the Nation 2025 report confirms that family background remains a strong predictor of adult earnings, with young people from higher professional households earning significantly more per hour than peers with identical qualifications (Social Mobility Commission, 2025). However, for an often-overlooked minority, that institution is absent or severely compromised. An estimated 16% of families in the UK are single-parent households (ONS, 2026), around 190,000 children have an incarcerated parent (Ministry of Justice, 2024), and hundreds of thousands more grow up in households affected by parental alcohol dependency, mental illness, or entry into state care. What these circumstances share is not merely financial hardship, but the erosion of a relational infrastructure: the shared meals, the school-day conversations, the parent-child relationship that transmits values, motivation, and aspirational norms.

This dissertation investigates the impact of family cohesion at age 14 on gross weekly wages at age 32, and whether the direction of this impact is conditional on the socioeconomic status of the parent.

The empirical analysis draws on the Next Steps' national cohort study of 15,770 individuals in England, which started whilst participants were aged 14 in 2004 and followed through to age 32. This dissertation begins by using Principal Axis Factoring to produce a Family Cohesion Factor (FCF) based on validated dimensions in the literature (encompassing the quality of parent-child relationships, meaningful communication, shared activities, and the absence of conflict) which is then used in an OLS log-wage regression. The credibility of the headline result is established through a sequence of increasingly stringent specifications and robustness checks. A Weighted Least Squares (WLS) model corrects for non-random attrition; a Heckman two-step correction provides directional evidence on selection into employment; a school fixed-effects across 608 schools absorbs time-invariant unobservables; Oster Bounds formally quantifies sensitivity to omitted variable bias; and a mediation analysis attempts to identify the specific pathway through which family cohesion may impact future adult earnings.

2. Literature Review

The dissertation bases its foundational economic theory on Becker's (1962) explanation of human capital investment. He defines human capital as the stock of knowledge, health, and values embedded within an individual to raise their productive capacity and earning potential. Importantly, Becker further specifies that factors such as emotional health, motivation, and knowledge “about the economic system”, like industry wage differentials directly impact the potential earnings of an individual. Becker and Tomes (1986) formalise this into an earnings model where cultural endowments act as a multiplier for financial investments in human capital. An adolescent who has grown up in an environment where discipline, punctuality, and occupational aspirations are the norm will extract more value from formal tuition than one who has not.

As Coleman (1988) outlines, the family unit serves as a reservoir of potential, containing both material assets (e.g. funding for study materials) and intellectual assets (e.g. parent's professional knowledge). However, these resources remain dormant without family cohesion. Coleman argues that capital within a family is functionally irrelevant to a child's development if the social relations required to transmit them are absent. This “social capital” is activated through everyday interactions, such as shared meals or discussions about school, which transform existing family structures into a mechanism for development. However, Granovetter's (1973) earlier theory, “strength of weak ties”, provides nuance to the topic: whilst strong family cohesion facilitates the transfer of values, it can be a closed loop. Economic mobility often relies on “weak ties” (acquaintances or distant professional contacts) who can provide market intelligence (e.g. unlisted vacancy) or social leverage like a strategic referral.

Nonetheless, the role of family cohesion in fostering non-cognitive capability provides advantages in the labour market. Heckman and Rubinstein (2001) demonstrate through a longitudinal study of US high-school dropouts that traits like persistence, reliability, and self-discipline are statistically significant determinants of earnings, independent of cognitive ability. They explicitly link the formation of these traits to the developmental home environment, arguing that motivation and self-discipline are “malleable” and fostered by parental encouragement and discipline. Kim and Torquati (2021) extend this to parent-child closeness specifically, finding that cohesion with parents predicted the development of responsibility, which in turn improved financial management behaviours. Leppard and Dufur (2025) provide contemporaneous evidence, demonstrating with US longitudinal data that

adolescent family social capital (operationalised through shared activities and parent-child communication) significantly increases the likelihood of successfully matching career goals a decade later. However, the literature stops short in estimating an economic market value of those achievements. The most proximate empirical contribution to the central question of this dissertation is provided by Millemaci and Sciulli (2014) who analyse the 1958 British National Child Development Study. Their results indicate that family difficulties reported at age 7 are associated with persistent labour market penalties extending up to the age of 51, providing strong longitudinal evidence of long-run economic associations.

Critically, the literature demonstrates that the incentives cohesion fosters are context dependent. Within high-income countries, Daniele and Geys (2016) empirically show that strong family ties impede geographical mobility and labour force participation because they create a “closed” social network where an individual’s economic and social needs are met entirely within the family unit, which reduces the incentive to establish external contacts required to access broader opportunities. Conversely, in low-income extended families, Di Falco and Bulte (2015) find that the obligation to share future wealth within a cohesive family act as a disincentive to individual human capital investment, creating a free-rider effect. These contrasting findings suggest that while cohesion can enhance individual capabilities, its economic returns depend on the broader attitudes and environment.

A further source of divergence in the literature arises from measurement challenges, there is an absence of a single, agreed-upon definition of “family cohesion” within the literature, and studies sometimes operationalise the concept through proxies like single-parent status or the number of siblings (Coleman, 1988). This dissertation however, uses the methodological convergence of the most established instruments like the Family Adaptability and Cohesion Evaluation Scale IV (FACES IV; Olson, 2011), the McMaster Family Assessment Device General functioning subscale (FAD-GF12; Epstein, Baldwin and Bishop, 1983), and the Family Functioning and Cohesion Scale (FFCS; Tsoutsi and Dikeos, 2024). While differing in structure, these measures converge in their reliance on self-reported responses on core dimensions such as emotional bonding, affective involvement, and interpersonal support. Building on this shared foundation, this dissertation constructs a composite measure of family cohesion using Principal Axis Factoring (PAF), consistent with the factor-analytic approach in related scale developments (Olson, 2011; Tsoutsi and Dikeos, 2024).

Overall, despite variations in context and measurement, the literature converges on two central propositions: first, that family cohesion plays a formative role in shaping non-

cognitive and motivational endowments that influence earning potential; and second, that these associations persist over the life course. This dissertation extends the literature by quantifying the economic returns to family cohesion and by opening the “black box” of transmission through mediation analysis, identifying the pathways through which early family environments translate into adult earning outcomes.

The econometric frameworks underpinning the analytical choices of this dissertation are embedded in further detail within the methodology.

3. Data and Sampling Methodology

The data is drawn from **Next Steps** (formerly the Longitudinal Study of Young People in England), a national cohort study managed by UCL’s Centre for Longitudinal Studies and initiated by the Department for Education in 2004 with 15,770 participants aged around 14 from England. Early sweeps collected data from both cohort members and parents on family stability, socioeconomic background, and educational attitudes; later sweeps shifted to labour market outcomes, housing, and health as the cohort matured. The dataset was selected for its large sample size and the breadth of variables spanning adolescent family environment and adult economic outcomes (University College London, 2026).

3.1 Analytic Samples

The **main analytic sample** links Sweep 1 (age 14, 2004) to Sweep 9 (age 32, 2022) via the ‘NSID’ identifier. This sample was restricted to observations with non-missing values for the dependent, independent, and control variables. See Table 1 for full list. **N = 4,041**

The **Heckman selection sample** expands the main sample to include non-employed cohort members at Sweep 9 with the same baseline controls and an additional variable for the number of children during Sweep 9 (exclusion restriction). **First-stage: N = 5,760**

The **school fixed-effects (FE) sample** is a subset of the main sample. As within-school variance cannot be estimated from a single observation, singleton observations were removed. **N = 4,005 across 608 schools.**

The **Mediation sample** required additional valid variables for the mediator at age 16. **N = 3,750**

3.2 Variables

Table 1 - List of all variables

	Variable	Age 14 (Sweep 1)	Age 32 (Sweep 9)
Dependent Variable	Ln(Wage)		✓
Key Independent Variable	Family Cohesion Factor (FCF) ¹	✓	
Demographic Controls	Male (=1)	—	—
	Ethnic Minority (=1)	—	—
Family Background Controls	Parent SES: High (ref. Low) ²	✓	
	Parent SES: Intermediate (ref. Low)	✓	
	Single Parent Household at Age 14 (=1)	✓	
	No. of Siblings at Age 14	✓	
Labour Market Controls	Weekly Hours Worked (centred)		✓
	Weekly Hours Above 37.5 (FT threshold spline)		✓
Human Capital Controls	No Long-term Illness (=1)		✓
	Passed GCSE (=1)	—	—
Psychological Trait (Big Five) Controls	Neuroticism		✓
	Conscientiousness		✓
	Extraversion		✓
	Agreeableness		✓
	Openness		✓
Geographic Controls	Region: London (ref. England)		✓
	Region: Other (Wales / Scotland / N. Ireland)		✓

¹ Variables used in factor listed in Table 6

² Highest National Statistics Socio-economic Classification of either parent at Sweep 1 (age 14) based on the ONS' three-class version: A) Higher managerial, administrative and professional occupations. B) Intermediate occupations C) Routine and manual occupations (reference category)

Dependent variable: Log gross weekly earnings (£) at age 32. This captures total remuneration prior to deductions, such as tax, pension contributions, and student loan repayments.

Independent variable: The Family Cohesion Factor (FCF); construction detailed in Section 4.1

Control variables: To isolate the impact of family cohesion during youth on future adulthood earnings and to minimise omitted variable bias, several controls that are popular in existing literature have been included: demographic, labour market, geographic, socioeconomic background etc.

Notably, the list of controls also includes the number of siblings (Sweep 1) based upon Resource Dilution Theory (Downey, 1995), which posits that parental time, emotional capital, and financial resources are finite; as the number of siblings increases, the resource available to foster cohesion and support for a young person is impacted. The models also attempt to control for inherent individual abilities that may influence future labour market success. Hence, a binary indicator for achieving a passing grade at GCSE (a mandatory UK examination) is included. Controls for optional qualifications, such as university degree attainment and specific industry of employment at Sweep 9 (age 32) have been **explicitly excluded**. Including these would likely introduce mediation bias, as family background may indirectly increase wages through mechanisms like encouraging higher educational attainment or entry into specific professions.

Finally, the models control for non-cognitive traits using the OCEAN Big Five personality tests (Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism), measured as continuous scores at Sweep 9 (age 32). The literature demonstrates an association between non-cognitive traits and labour market outcomes (Heckman, Stixrud and Urzua, 2006). Although these scores were recorded in adulthood, psychological literature demonstrates that these core personality traits are remarkably stable across the life course (Roberts and DelVecchio, 2000) and behaviour genetics indicates that these traits are highly heritable (Jang, Livesley and Vernon, 1996). Therefore, these controls can act as a reasonable proxy for unobserved genetic personality traits of the parents, which heavily form the baseline emotional cohesion of the household during youth.

3.3 Summary statistics

Table 2 - Continuous variables

Variable	Mean	SD	Min	Max
Family Cohesion Factor	-0.01	0.79	-3.60	1.46
Gross Weekly Wage (£)	848.30	2,160.09	11.63	79,259.62
Log Gross Weekly Wage	6.43	0.69	2.45	11.28
Weekly Hours Worked	36.23	9.14	1	98
Big Five: Neuroticism	12.92	2.56	3	21
Big Five: Conscientiousness	14.29	2.15	3	21
Big Five: Extraversion	13.02	2.53	3	21
Big Five: Agreeableness	12.66	2.20	3	21
Big Five: Openness	13.24	3.51	3	21
Number of Siblings at Age 14	1.51	1.08	0	9
Number of Children at Age 32	0.78	0.97	0	10

Table 3 - Binary variables

Variable	= 1/Yes (%)	= 0/No (%)
Male	44	56
No Long-Term Illness	75	24
Passed GCSEs	78	22
Single-Parent Household at Age 14	18	82
Ethnic Minority	23	77

Table 4 - Categorical variables

Variable	Category	Proportion (%)
Parental SES	High	50
	Intermediate	30
	Low	20
Region	England (excl. London) ³	84
	London	15
	Other (Wales / Scotland / N. Ireland)	1

³ Next Steps sampled from schools exclusively in England (Sweep 1)

4. Methodology

4.1 Construction of the Family Cohesion Factor (FCF)

Within the constraints of secondary birth-cohort data, a 6-item youth-reported factor was constructed using the full Sweep 1 sample of valid respondents to the 6-items (N = 13,385). The variables were selected based on availability in Next Steps Sweep 1, and face-validity match with the cohesion subscales of FACES IV, FAD-GF12, and FFCS.

The factor was generated through four steps:

1. Variables were reverse-coded where necessary so that higher values consistently represent greater cohesion. Responses indicating an absent parent were coded as missing to preserve scale validity.
2. Items with differing Likert scales (3 to 5-point) were standardised to z-scores to ensure equal variance prior to aggregation
3. Where questions asked the young person to evaluate both parents separately, a "weakest link" assumption was applied based on family systems theory: the most conflict-heavy relationship in the household determines overall cohesion. Averaging two discrepant scores produces a value that neither is held by either participant nor captures the overall functional quality of the family environment (Cox and Paley, 1997). For single-parent households, or instances where a youth only reported on one parent, the metric defaults to the sole available parent's score. The designation of weakest link varies by dimension, fathers are identified as the weakest link in conflict (58%), while mothers dominate in relationship quality (87%) and communication (67%), a pattern partly driven by the higher prevalence of mother-only observations in the sample.

Table 5 - Breakdown of "weakest link" by parent

	Mother as "weakest link" (%)	Father as "weakest link" (%)
Relationship Quality	87	13
Meaningful Communication	67	33
Conflict	42	58

4. Principal Axis Factoring was applied to extract a single continuous latent factor. The KMO measure of sampling adequacy was 0.70 and Bartlett's test of sphericity was highly significant ($p < 0.001$), confirming the items were suitably intercorrelated for factor

analysis. The resulting factor demonstrated acceptable internal consistency for a 6-item scale (polychoric alpha = 0.70). Polychoric alpha was preferred to Cronbach's alpha for ordinal items, as Cronbach assumes interval-level measurement which Likert items violate (Gadermann, Guhn and Zumbo, 2012).

Table 6 - Polychoric alpha of FCF

Polychoric Alpha = 0.70				
	Mean ± SD*	Item-Total Correlation	Alpha if deleted	Factor Loading
Total score	-0.09 ± 0.61			
1. Quality of relationship with mother / father	-0.18 ± 1.10	0.73	0.60	0.79
2. Communication with mother / father about important matters	-0.23 ± 0.99	0.62	0.63	0.63
3. Frequency of conflicts with mother / father	-0.20 ± 1.03	0.55	0.66	0.57
4. Parental communication about school day	0.01 ± 1.00	0.50	0.66	0.47
5. Frequency of family evening meals	0.01 ± 0.99	0.38	0.69	0.37
6. Frequency of future study plan discussions with family	0.04 ± 0.98	0.35	0.70	0.33

KMO = 0.70, Bartlett p-value < 0.001

4.2 Baseline Specification

The initial empirical strategy begins with an OLS regression to establish the baseline relationship between the Family Cohesion Factor (FCF) and adult wages. Following standard practice, a natural log is applied to the wage variable. This transformation corrects severe positive skewness (from 24.81 to 0.20) and kurtosis (from 742.42 to 4.17). Despite slight remaining kurtosis, the large samples ($N > 4,000$) ensures valid OLS inference via the Central Limit Theorem.

The baseline takes the form:

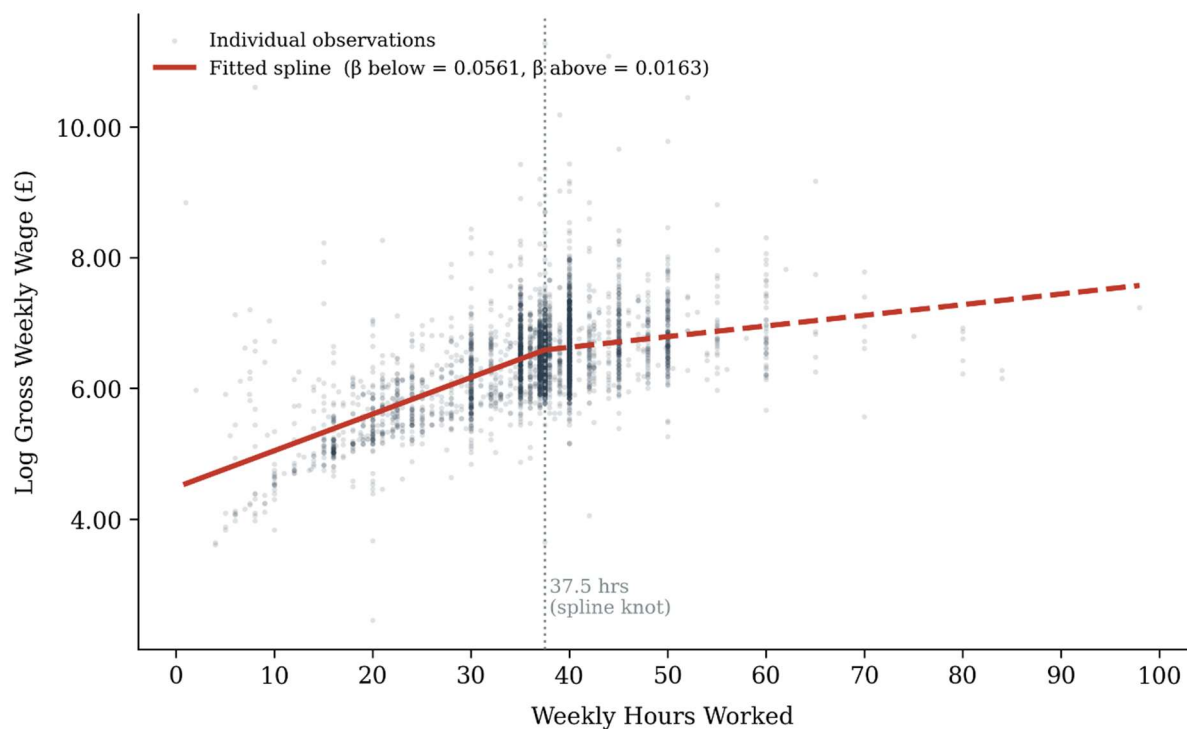
$$\ln(Wage) = \alpha + \beta_1(FCF) + X'\gamma + \varepsilon_i$$

Where $\ln(Wage)$ is the natural logarithm of gross weekly wages at age 32, β_1 is the coefficient of the primary variable of interest such that a one standard deviation increase in

the FCF, measured at age 14, is associated with an approximate $100 \times \beta_1$ percentage change in earnings. X' is the vector of control variables described in Section 3.2, and ε_i is the error term.

To isolate the earning capacity effect of family cohesion from mere hours worked, all specifications include a control for weekly hours worked. Given that the returns to an additional hour of work are unlikely to be constant across part-time and full-time margins (Moffitt, 1984), hours are entered as a piecewise linear spline with a knot at 37.5 hours per week (a common contractual amount). As shown in Figure 1, the scatterplot of log wages against hours reveals a pronounced change in gradient within the 37.5-hour region. By permitting the regression slope on hours to differ on either side of this threshold, the models capture the distinct wage-hours relationship (Manning and Petrongolo, 2008).

Figure 1 - Hours Worked Vs. Wages



4.3 Survey Context, Weights, and Complex Design

The nature of the Next Steps study introduces specific biases that this methodology seeks to mitigate. First, the study used a complex, stratified sampling design. Cohort members were drawn from schools acting as Primary Sampling Units (PSUs), with a deliberate over-sampling of deprived schools and ethnic minority groups to ensure longitudinal analysis

remains representative in the long term. While necessary for depth, this design requires weighting to correct for unequal selection probabilities and to prevent bias in descriptive or causal estimates (Solon, Haider and Wooldridge, 2013).

Second, the study has faced attrition over its 18-year span. Of the 15,770 cohort members in 2004, only 7,279 (46%) participated in Sweep 9. As documented in the study's technical guide, this attrition was non-random, resulting in a disproportionate loss of certain demographics that could undermine the validity of unweighted findings.

To address this, the study provides a specific longitudinal weight (W9FINWT) constructed using random forest modelling that drew upon around 50 predictor variables (Adali *et al.*, 2025). Crucially, the weight construction also included predictors across family structure, sibling count, and parental SES. The documentation confirms that this weighting successfully corrects for non-random attrition bias across these specific predictors. Since the weights incorporate the same family background dimensions central to this dissertation, the resulting weighted estimates effectively correct for attrition bias along those specific dimensions.

In practice, because the W9FINWT variable is scaled to the full Sweep 9 sample size ($N = 7,279$), this methodology re-normalises weights for each specific analytical sub-sample in accordance with the literature (Solon, Haider and Wooldridge, 2013).

4.4 Heckman Correction

A fundamental challenge in wage regressions is sample selection bias.

To address this, a Heckman Two-step correction is employed.

$$P(\text{Employed} = 1 | X) = \Phi(\alpha + \beta_1(\text{FCF}) + \dots \beta_k(\text{No. of Children at Age 32}))$$

In the first stage, a survey-weighted probit model estimates the probability of being employed at Sweep 9 (Age 32) using a sample comprised of both employed and non-employed participants. Following the precedent set by Heckman (1974) and the subsequent literature, the number of children a cohort member has at age 32 is used as an exclusion restriction. This variable is theoretically linked to labour market participation (particularly for women) but is assumed to have no direct effect on the gross earnings capacity outside of labour supply decisions.

From this first stage, an Inverse Mills Ratio (IMR), denoted as λ , is computed to represent the relative probability of an individual being in the sample:

$$\lambda = \frac{\phi(X)}{\Phi(X)}$$

Where ϕ and Φ are the standard normal probability density and cumulative distribution functions. The IMR is then included as an additional control variable in the second-stage WLS equation to consistently estimate the parameters while adjusting for selection bias.

$$\ln(Wage) = \alpha + \beta_1(FCF) + \beta_k(\lambda) + X'\gamma + \varepsilon_i$$

4.5 School Fixed-Effects (FE)

A school FE is estimated to further address concerns about unobserved school and neighbourhood characteristics, such as school quality which has shown positive effects on adult earnings in the literature (Altonji and Mansfield, 2018), or neighbourhood-level characteristics, like community aspirations and the socioeconomic composition of the catchment area which exert long-run outcomes (Chetty, Hendren and Katz, 2016). Furthermore, schools may generate social capital through peer groups and alumni networks in ways unobservable within the dataset.

The FE specification is operationalised by including a full set of Primary Sampling Unit (School) dummies, absorbing all time-invariant, school and neighbourhood-level unobservables, thereby restricting variation in the FCF-wage relationship to that within schools.

4.6 Instrument Variable (IV)

To address potential endogeneity, two instruments were pursued in line with the literature and data availability. The first was parental religiosity at Sweep 1, which is shown to predict emotional bonding (Mahoney, 2010); the second was first-born status, as first-born children receive significantly greater parental quality time than later-born siblings (Price, 2008). In practice, neither instrument satisfied the relevance condition: first stage excluded-instrument F-statistics were 0.17 and 0.39 respectively, far below the conventional threshold of 10.

IV estimation was therefore not pursued further.

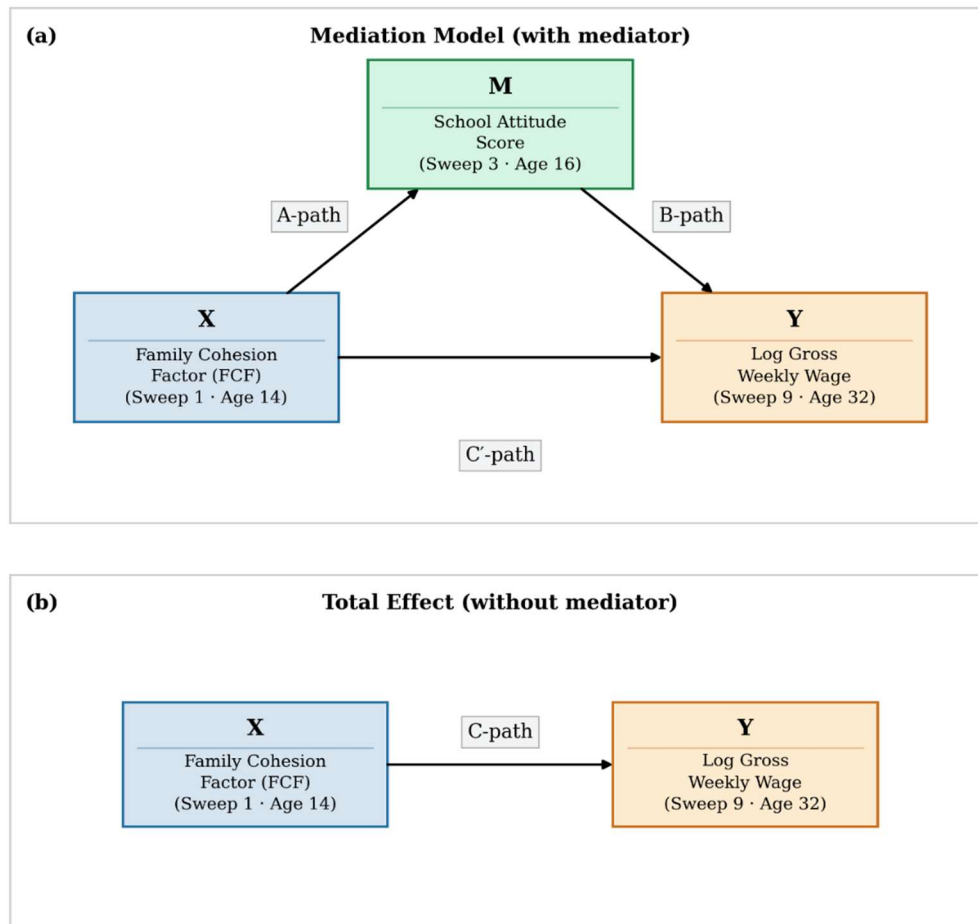
4.7 Mediation Analysis

To investigate whether the relationship between family cohesion X and adult wages Y operates through a mediator M , this study decomposes the relationship into components.

As illustrated in Figure 2's panel (a), the total effect of X and Y is separated into an indirect effect which transmits through an intermediate variable M (A-path to B-path). In addition, to

a direct effect, which captures the residual association between X and Y once M is held constant (C'-path).

Figure 2 - Mediation Model



Multiple potential mediators were tested in line with the literature, such as mental health scores at age 17 (Fletcher, 2013) and juvenile delinquency (Dorsett and Thomson, 2025). However, these variables failed to yield statistically significant mediating pathways. Therefore, in the present analysis, the mediator selected is the cohort member's School Attitude Score (0-48) at age 16 (Sweep 3). This additive measure is derived from twelve questionnaire items assessing enjoyment, engagement, and perceived value of school⁴ (Next Steps, 2006). This pathway is theoretically motivated by the literature which demonstrates that school engagement at age 16 has a persistent effect on incomes during adulthood (Symonds, D'Urso and Schoon, 2023). The temporal sequencing of the three variables across Sweep 1 (Age 14), Sweep 3 (Age 16), and Sweep 9 (Age 32) provides a natural ordered

⁴ See Annex, Table 18 for the full components/description of the additive variable. The variable was sourced directly from Next Steps, and no further manipulation was performed.

structure that is consistent with the causal direction assumed by the mediation framework. The mediation model is analysed through three OLS equations corresponding to the paths in Figure 2:

A-path: $School\ Attitude = Constant + A(FCF) + X'\gamma + \varepsilon_i$

C'-path: $\ln(Wage) = Constant + C'(FCF) + B(School\ Attitude) + X'\gamma + \varepsilon_i$

C-path: $\ln(Wage) = Constant + C(FCF) + X'\gamma + \varepsilon_i$

Notably, the OLS regressions in the mediation analysis have a few intended structural differences compared to the baseline specification. First, personality traits and the “passed GCSE” variables have been excluded from the mediation models, as they are measured/achieved contemporaneously or after the mediator (whilst also being variables that could constitute downstream mediators) thus potentially inducing bias. Second, the sample has been restricted to those in employment and have complete data on all three pathway variables leading to all three equations having $N = 3,750$.

5. Results

5.1 Central Result

The headline result is consistent across all specifications. A one standard deviation rise in the Family Cohesion Factor (FCF) at age 14 is associated with approximately 2.2 to 3.3% higher gross weekly wages at age 32, and this association is statistically significant at $p < 0.05$ across all specifications, except Heckman S2 with marginal significance at $p < 0.1$.

Table 7 - Central Specifications

$p < 0.1$ *, $p < 0.05$ **, $p < 0.01$ ***

Dependent variable: Ln(Wage)

	(1)	(2)	(4)	(5)
	OLS	WLS	Heckman S2	School FE
Constant	6.469*** (0.104)	6.365*** (0.117)	6.551*** (0.152)	— —
Family Cohesion Factor (FCF)	0.029*** (0.010)	0.028** (0.011)	0.022* (0.012)	0.033** (0.013)
Male (=1)	0.081*** (0.019)	0.074*** (0.023)	0.056** (0.024)	0.090*** (0.025)
Ethnic Minority (=1)	0.056** (0.024)	0.049 (0.030)	0.048 (0.030)	0.068* (0.039)
Parent SES: High (ref. Low)	0.197*** (0.024)	0.160*** (0.030)	0.119*** (0.036)	0.089*** (0.031)
Parent SES: Intermediate (ref. Low)	0.078*** (0.025)	0.027 (0.030)	-0.001 (0.034)	0.035 (0.032)
Single Parent Household at Age 14 (=1)	-0.046** (0.021)	-0.056** (0.026)	-0.037 (0.027)	-0.041 (0.028)
No. of Siblings at Age 14	-0.018** (0.008)	-0.008 (0.009)	0.000 (0.010)	-0.006 (0.011)
Labour Market Controls	✓	✓	✓	✓
Human Capital Controls	✓	✓	✓	✓
Psychological Trait (Big Five) Controls	✓	✓	✓	✓
Geographic Controls	✓	✓	✓	✓
Inverse Mills Ratio (λ)	—	—	-0.272** (0.132)	—
Observations (N)	4041	4041	4041	4005
No. of School FEs	—	—	—	608
Adj. R²	0.404	0.428	0.429	—
F-statistic	153.35	168.90	160.43	—
Adj. R² (within)	—	—	—	0.398
F-statistic (on covariates, within)	—	—	—	7.30

HC3 heteroskedasticity-robust standard-error. WLS, Heckman S2, and School FE weighted using W9FINWT. Model (3) is the Heckman first-stage within the Annex.

The WLS estimate, which adjusts the sample composition to better represent the original cohort, yields a slightly attenuated coefficient compared to the OLS baseline (2.8% compared to 2.9%); this is consistent with mild positive selection in the unweighted analytic sample.

The first-stage employment probit⁵ exhibits an acceptable fit (McFadden Pseudo R-squared = 0.184, Wald $p < 0.001$), the exclusion restriction performs as expected ($\beta = -0.159$, $p < 0.001$), and the FCF coefficient in stage-1 is positive and significant ($\beta = 0.083$, $p < 0.01$), indicating that family cohesion **operates on the participation margin as well as the wage margin**. This is a dimension the other specifications cannot capture due to being estimated on the employed sub-sample. The second-stage FCF coefficient (2.2%) broadly corroborates the other specifications, though sits at the lower end of the band.

The IMR is negative and significant ($\lambda = -0.272$, $p < 0.05$); this is consistent with labour-supply literature on the negative selection into employment on wages (Killingsworth and Heckman, 1986). The Jarque-Bera test rejects bivariate normality ($p < 0.001$), so the exact second-stage magnitudes are approximate. The Heckman specification is therefore retained as suggestive evidence on selection, not as refined point estimates.

Finally, the narrative was strengthened by arguably the most conservative specification. A school fixed-effects (FE) model estimated across 608 schools compared cohort members who attended the same institution in 2004, absorbing time-invariant differences like school quality, community aspirations, and neighbourhood characteristics. This specification yielded a statistically significant estimate coefficient of 0.033 ($p < 0.05$) and at the highest range of the headline models.

5.2 Diagnostic Tests and Standard Error Strategy

As expected in wage regressions, the Breusch-Pagan test returned p-values below 0.001 across all specifications, indicating the presence of heteroskedasticity. HC3 heteroskedasticity-robust standard-errors were therefore applied throughout (Long and Ervin, 2000), in preference to clustering at the school level, because the variable of interest (FCF) varies at the individual household level, and the Next Steps cohort excludes siblings/twins, so there is no meaningful within-school residual correlation attributable to the treatment variable itself.

⁵ Table 15 shows the full results

However, Abadie *et al.* (2022) argue that clustering could be warranted whenever the sample is drawn in a clustered manner, since observations within the same cluster shared unobserved characteristics that can induce residual correlation. As a robustness check, school-level clustered standard errors were estimated on the school FE model. The differences were negligible: HC3 standard errors were on average +0.0019 larger than their clustered counterparts, meaning HC3 is the more conservative choice and is therefore used consistently across all specifications.

5.3 Sensitivity to Omitted Variables

Despite the rich set of controls, it is reasonable to expect some omitted variables (e.g. parenting style (Doepke and Zilibotti, 2017)) influencing the FCF-wage relationship. To formally quantify the degree to which unobservable variables could overturn the headline result, an Oster (2019) bound test was implemented. This approach exploits the joint movement of the coefficient and R-squared as controls are sequentially added to bound the bias-adjusted treatment effect (β^*) and compute δ , which is the ratio of selection on unobservables to selection on observables required to drive the true effect to zero.

Table 8 - Oster Bounds

Model	Baseline Effect (BE)		Controlled Effect (CE)		Bias-Adjusted $R_{max} = 1.3\tilde{R}$		Bias-Adjusted $R_{max} = 0.90$	
	β_{BE}	R^2_{BE}	β_{CE}	R^2_{CE}	β^*	δ	β^*	δ
OLS (1)	0.042** (0.013)	0.003	0.029*** (0.010)	0.407	0.025	6.851	0.012	1.710
WLS (2)	0.051*** (0.015)	0.004	0.028** (0.011)	0.430	0.021	3.785	0.001	1.048

Results are reported under two assumptions for R_{max} (the r-squared of a hypothetical regression with **no omitted variables**). The default assumption is that of 1.3x the original r-squared based on Oster's recommendation, and a more conservative stress test of 0.90 has also been implemented.

Under $R_{max} = 1.3\tilde{R}$ both specifications are robust. The OLS bias-adjusted coefficient remains positive ($\beta^* = 0.025$) with $\delta = 6.851$, and the WLS yields ($\beta^* = 0.028$) with $\delta = 3.785$, indicating that unobservables would need to be nearly four to seven times more

influential than all the currently included controls to nullify the effect. Even whilst stress-testing to a hypothetical $R_{max} = 0.90$, both models remain robust at the conventional threshold of $\delta > 1$.

5.4 Non-Linearity

The literature posits that the relationship between family cohesion and developmental outcomes is non-linear. Both insufficient cohesion (disengagement) and excessive cohesion (enmeshment) are dysfunctional (Olson, 2011), providing a theoretical basis for testing non-linear specification.

A quadratic term (FCF^2) was added to both stages of the Heckman correction model. The results diverge meaningfully across the two stages.

Table 9 - FCF Quadratic specification

Stage	$\beta_1 (FCF)$	$\beta_2 (FCF^2)$	Turning Point
Employment Probit	0.033 (0.046)	-0.042* (0.025)	68.4 th percentile
Ln(Wage)	0.034** (0.015)	0.012 (0.009)	6.3 rd percentile

In the first-stage employment probit, the squared FCF term is negative and marginally statistically significant ($p < 0.1$) with an implied turning point at the 68.4th percentile of the FCF distribution, meaning beyond that threshold, additional cohesion is associated with a declining probability of labour market participation. This coincides with the literature which indicates those who were nurtured in overly cohesive or “helicopter parenting” households exhibit greater career self-doubt, reduced adaptability, and lower career exploration (LeBlanc and Lyons, 2022). However, **the linear FCF term in this stage is not significant**, so this is best read as suggestive of the shape at the upper tail, not a precise threshold.

On the other hand, the second-stage wage equation’s squared FCF term is positive but statistically insignificant with an effectively extreme turning point at the 6.3rd percentile. The linear FCF term remains significant and close to the baseline estimate, suggesting that the analytical sample is well-approximated by a linear function.

5.5 Robustness of the FCF construct and Gender/Socioeconomic Heterogeneity

The FCF was constructed under a “weakest link” assumption grounded in family systems theory, whereby the more conflicted parental relationship determines the overall cohesion score. To assess whether this choice drives the headline findings, four alternative FCF specifications were estimated within the WLS framework: a simple average of both parental scores, a “strongest link” model using the higher-scoring parent, and a dedicated mother-focused and father-focused specifications.

Table 10 - FCF variants

$p < 0.1$ *, $p < 0.05$ **, $p < 0.01$ ***

Dependent variable: Ln(Wage)					
	(2)	(2A)	(2B)	(2C)	(2D)
FCF Specification:	Weakest Link	Average	Strongest Link	Mother-focused	Father-focused
Family Cohesion Factor (FCF)	0.028** (0.011)	0.033*** (0.013)	0.034*** (0.013)	0.045*** (0.015)	0.045** (0.019)
Male (=1)	0.074*** (0.023)	0.075*** (0.023)	0.076*** (0.023)	0.071*** (0.023)	0.068*** (0.026)
FCF x Male	—	—	—	-0.041* (0.024)	-0.013 (0.027)
Default Model Controls	✓	✓	✓	✓	✓
Observations (N)	4041	4041	4041	3932	3254
Adj. R ²	0.428	0.429	0.429	0.425	0.389
F-statistic	168.90	169.00	168.94	153.61	110.10

HC3 heteroskedasticity-robust standard-error. Weighted using W9FINWT.

The weakest link, average, and strongest link specifications yield significant FCF coefficients of 0.028, 0.033, and 0.034 respectively. This narrow band demonstrates that the headline figure is unlikely to be an artefact of the aggregation method. The marginal increase from weakest to strongest link indicates that a child’s development is most effectively shaped by their most functional parental relationship.

Nevertheless, the weakest link assumption is retained as preferred: the empirical differences are small, and the specification is motivated by the well-established family systems literature.

The parent-specific models yield uniformly larger coefficients (0.045 in both the mother and father-focused specifications), a pattern explained by measurement attenuation in the aggregated models. The smaller father-focused sample (N = 3,254) relative to the mother-focused sample (N = 3,932) reflects two partially overlapping sources of parental data sparsity in Sweep 1. The first is genuine parental absence from a cohort member's life, which accounts for 18% of the analytical sample. However, there is a second minority scenario where fathers in otherwise intact dual-parent households did not participate in the study. This dissertation defaults to the sole available parent in cases of unilateral responses, because listwise deletion would systematically exclude observations most likely to involve lower parental engagement, introducing its own selection bias into the FCF construction.

Table 11 - Gender Interaction Specification

$p < 0.1$ *, $p < 0.05$ **, $p < 0.01$ ***

Dependent variable: Ln(Wage)				
	(6)	(7)	(9)	(10)
	OLS	WLS	Heckman S2	School FE
Family Cohesion Factor (FCF)	0.034*** (0.013)	0.044*** (0.015)	0.038** (0.015)	0.044*** (0.017)
Male (=1)	0.079*** (0.019)	0.067*** (0.024)	0.049** (0.025)	0.085*** (0.025)
FCF x Male	-0.012 (0.020)	-0.035 (0.022)	-0.036 (0.022)	-0.025 (0.026)
Default Model Controls	✓	✓	✓	✓
Observations (N)	4041	4041	4041	4005
No. of School FEs	—	—	—	608
Adj. R ²	0.404	0.428	0.429	—
F-statistic	145.28	160.28	152.62	—
Adj. R ² (within)	—	—	—	0.398
F-statistic (on covariates, within)	—	—	—	7.29

The specifications include an interaction term; therefore, the main FCF coefficient identifies the female-specific premium and is NOT directly comparable to the pooled estimates in Table 7. HC3 heteroskedasticity-robust standard-error. WLS, Heckman S2, and School FE weighted using W9FINWT.

Turning to gender heterogeneity, the FCF x Male interaction is negative across all four main specifications. However, a subtle pattern emerges in the parent-focused robustness check. The

mother-focused variant produces a marginally significant and negative interaction (-0.041 , $p < 0.1$), implying a near-zero cohesion premium for male cohort members via the maternal relationship ($0.045 - 0.041 = 0.004$), against a female premium of 0.045 . This mother-daughter dynamic aligns with existing literature, which finds labour market outcomes to be strongest shaped through the mother-daughter role-model mechanism, in contrast to all other pairings (Galassi, Koll and Mayr, 2024).

The FCF interactions with both High and Intermediate parental SES are statistically insignificant across all four main specifications⁶. The complete absence of significance at any level, across any model (even Heckman Stage-1)⁷ provides clean null evidence. The earnings premium to youth family cohesion does not vary by parental socioeconomic class.

5.6 Mediation Analysis

The A-path estimate confirms that family cohesion is a strong positive predictor of school attitude at age 16. A one standard deviation increase in the FCF at age 14 is associated with a 0.266 standard deviation increase in the school attitude score at age 16 ($p < 0.001$), aligning with the literature's assertion on the family environment's positive impact on academic success (Zhao and Zhao, 2022). Moreover, the B-path confirms that a one standard deviation increase in the school attitude score is associated with a 0.098 standard deviation increase in wages at age 32 ($p < 0.001$). Crucially, once school attitude is held constant in the C'-path specification, the direct effect of a one standard deviation increase in FCF is statistically insignificant on adult wages.

⁶ See Annex. Table 16

⁷ See Annex. Table 15

Table 12 - Mediation Specifications

$p < 0.1$ *, $p < 0.05$ **, $p < 0.01$ ***

Mediator: Attitude to School at Age 16			
	(16)	(17)	(18)
Dependent variable	Attitude to School Score	Ln(Wage)	Ln(Wage)
Path	$X \rightarrow M$ (A-path)	$M \rightarrow Y X$ (C'-path)	$X \rightarrow Y$ (C-path)
Constant	-0.126** (0.054)	-0.250*** (0.050)	-0.269*** (0.050)
Family Cohesion Factor (FCF)	0.266*** (0.016)	0.008 (0.014)	0.034** (0.014)
Attitude to School Score at Age 16	— —	0.098*** (0.014)	— —
Male (=1)	-0.086*** (0.032)	0.166*** (0.030)	0.153*** (0.030)
Ethnic Minority (=1)	0.201*** (0.039)	0.198*** (0.036)	0.216*** (0.036)
Parent SES: High (ref. Low)	0.207*** (0.046)	0.336*** (0.037)	0.356*** (0.038)
Parent SES: Intermediate (ref. Low)	0.065 (0.048)	0.125*** (0.038)	0.130*** (0.038)
Single Parent Household at Age 14 (=1)	-0.206*** (0.045)	-0.039 (0.032)	-0.059* (0.032)
No. of Siblings at Age 14	0.020 (0.016)	-0.037*** (0.013)	-0.035*** (0.013)
Weekly Hours Worked (centred)	— —	0.074*** (0.003)	0.075*** (0.003)
Weekly Hours Above 37.5 (FT threshold spline)	— —	-0.055*** (0.006)	-0.055*** (0.006)
No Long-term Illness (=1)	— —	0.118*** (0.030)	0.131*** (0.030)
Observations (N)	3750	3750	3750
Adj. R ²	0.088	0.369	0.360
F-statistic	52.48	200.03	211.96

The mediation decomposition in Table 13 reveals a positive and significant indirect effect, with the 95% confidence interval entirely excluding zero. Notably, the total effect (0.034) is consistent with the baseline OLS coefficient (0.029 from Table 7). This suggests that the mediation sub-sample accurately reflects the FCF-wage gradient observed in the main sample, despite the differences (Section 4.7).

Table 13 - Mediation Effect Decomposition

	Estimate	95% Confidence Interval
Indirect Effect (via School Attitude Score)	0.026	0.018, 0.034
Direct Effect	0.008	-0.018, 0.035
Total Effect	0.034	0.010, 0.059
Bootstrap Replications	1,000 replications (1,000 converged)	

Green-shading = CI excludes zero. Bootstrap 95% CI method: percentile (cluster PSU).

The proportion of the total effect transmitted via school attitude is approximately 76% (0.026/0.034). However, ratio-based mediation metrics become unstable when the total effect is small (MacKinnon, 2012). It is important to also acknowledge that this mediation decomposition is descriptive of the statistical pathway within the data rather than a confirmed causal mechanism.

The results suggest a higher FCF at age 14 is positively associated with a stronger orientation towards school at age 16 which predicts a higher wage at age 32. This is consistent with the broader family investment model, whereby cohesive families are more likely to engage in educational discussions (captured by the FCF's 4th and 6th item), provide stable environments conducive to learning (Fosco and Lydon-Staley, 2020), and transmit aspirational norms that propel academic motivation (Wu *et al.*, 2018).

These findings do not conflict with the Oster bounds established in section 5.3, as the Oster test interrogated the robustness of the **total effect** (C-path), whereas the mediation analysis decomposes the pathways of this association.

6 Limitations

The absence of a valid IV leaves endogeneity formally unresolved. Unlike settings where simultaneity is a concern, the temporal ordering of this study (FCF measured at age 14, wages at age 32) reduces concerns about reverse causality. The residual threat is instead confounding, for instance, a child's innate temperament plausibly generates higher FCF through eliciting warmer parenting responses, and independently higher adult wages through labour market-relevant traits like resilience. The Oster Bounds partially addresses this by looking at the limits of how much unobservables could sway the effect.

Additionally, the FCF measure relies solely on youth-reported responses/perceptions and captures cohesion at a single point in time (age 14), potentially missing key life-changes in the subsequent years of adolescence.

More substantively, labour market experience which is a central variable in Mincer's (1974) wage equation is absent due to a structural gap in the Next Steps study. Between the ages of 20 and 32 only one sweep took place, therefore individual job histories were untracked.

7 Conclusions

This dissertation set out to investigate the impact of family cohesion at age 14 on gross weekly wages at age 32, and whether the direction of this impact is conditional on the socioeconomic status of the parent.

Across a sequence of increasingly stringent specifications, including WLS, Heckman selection corrections, and school fixed effects: a one standard deviation increase in the Family Cohesion Factor (FCF) measured at age 14 is consistently associated with a 2.2-3.3% premium in gross weekly wages at age 32. An Oster bound test indicated that unobservables would need to be between four to seven times more influential than the full set of included controls (human capital, labour supply, Big Five personality traits, and family background) to nullify the effect. However, the absence of a valid instrumental variable means the estimates are best interpreted as robust associations rather than a causal effect.

Additionally, the absence of any statistically significant interaction between FCF and parental SES at age 14 provides clean, null evidence regarding conditionality. However, the analysis also reveals important nuances within this overarching trend. The Heckman first-stage provides suggestive evidence that family cohesion operates on the participation margin as

well as the wage margin, with suggestive non-linearity at the upper tail. Furthermore, gender heterogeneity tests provide suggestive evidence of a stronger mother-daughter transmission, aligning with existing literature on same-sex role modelling.

The mediation analysis opened the “black box” of the FCF-wage transmission, revealing that a substantial portion of the wage premium associated with family cohesion operates indirectly through the cohort member’s attitude to school at age 16. This carries practical implications for policymakers and institutions. As the FCF-wage association appears to operate largely through a modifiable intermediate state (school engagement) there is an alternative policy lever compared to intrusive family intervention, such as targeted school pastoral care for those at risk of “low cohesion”.

While this dissertation establishes a robust narrative, several avenues for future research could be explored. For example, to further limit the endogeneity concerns, future research could benefit from a panel dataset that samples multiple children per household to allow for a sibling fixed effects approach that utilises differential experiences of cohesion within the same family to estimate cleaner impacts on earnings. Additionally, the current methodology interacted family cohesion with broad SES categories, which may be too crude to capture direct professional advantages. An alternative approach could be to track children who go onto work in the same industry or occupation as their parents whilst accounting for different levels of FCF to explicitly test Granovetter’s “strength of weak ties” theory.

Ultimately, extending this research across different cohorts and methodologies will refine the field’s understanding of how the family continues to shape economic outcomes many years later.

8 References

- Abadie, A. *et al.* (2022) ‘When Should You Adjust Standard Errors for Clustering?’, *The Quarterly Journal of Economics*, 138(1), pp. 1–35. Available at: <https://doi.org/10.1093/qje/qjac038>.
- Adali, T. *et al.* (2025) ‘Next Steps: Sweep 9 Age 32’. Available at: https://doc.ukdataservice.ac.uk/doc/5545/mrdoc/pdf/next_steps_sweep_9_age_32_user_guide_v2.pdf.
- Altonji, J.G. and Mansfield, R.K. (2018) ‘Estimating Group Effects Using Averages of Observables to Control for Sorting on Unobservables: School and Neighborhood Effects’, *American Economic Review*, 108(10), pp. 2902–2946. Available at: <https://doi.org/10.1257/aer.20141708>.

- Becker, G.S. (1962) 'Investment in Human Capital: A Theoretical Analysis', *Journal of Political Economy*, 70(5, Part 2), pp. 9–49. Available at: <https://doi.org/10.1086/258724>.
- Becker, G.S. and Tomes, N. (1986) 'Human Capital and the Rise and Fall of Families', *Journal of Labor Economics*, 4(3, Part 2), pp. S1–S39. Available at: <https://doi.org/10.1086/298118>.
- Chetty, R., Hendren, N. and Katz, L.F. (2016) 'The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment', *American Economic Review*, 106(4), pp. 855–902. Available at: <https://doi.org/10.1257/aer.20150572>.
- Coleman, J.S. (1988) 'Social Capital in the Creation of Human Capital', *American Journal of Sociology*, 94, pp. S95–S120. Available at: <https://doi.org/10.1086/228943>.
- Cox, M.J. and Paley, B. (1997) 'FAMILIES AS SYSTEMS', *Annual Review of Psychology*, 48(1), pp. 243–267. Available at: <https://doi.org/10.1146/annurev.psych.48.1.243>.
- Daniele, G. and Geys, B. (2016) 'Family Ties and Socio-Economic Outcomes in High vs Low Income Countries', *The Journal of Development Studies*, 52(6), pp. 813–823. Available at: <https://doi.org/10.1080/00220388.2015.1098630>.
- Di Falco, S. and Bulte, E. (2015) 'Does social capital affect investment in human capital? Family ties and schooling decisions', *Applied Economics*, 47(2), pp. 195–205. Available at: <https://doi.org/10.1080/00036846.2014.967383>.
- Doepke, M. and Zilibotti, F. (2017) 'Parenting With Style: Altruism and Paternalism in Intergenerational Preference Transmission', *Econometrica*, 85(5), pp. 1331–1371. Available at: <https://doi.org/10.3982/ECTA14634>.
- Dorsett, R. and Thomson, D. (2025) 'How Incarcerating Children Affects their Labour Market Outcomes', *Journal of Social Policy*, 54(2), pp. 465–486. Available at: <https://doi.org/10.1017/S0047279423000235>.
- Downey, D.B. (1995) 'When Bigger Is Not Better: Family Size, Parental Resources, and Children's Educational Performance', *American Sociological Review*, 60(5), p. 746. Available at: <https://doi.org/10.2307/2096320>.
- Epstein, N.B., Baldwin, L.M. and Bishop, D.S. (1983) 'The McMaster Family Assessment Device', *Journal of Marital and Family Therapy*, 9(2), pp. 171–180. Available at: <https://doi.org/10.1111/j.1752-0606.1983.tb01497.x>.
- Fletcher, J. (2013) 'Adolescent Depression and Adult Labor Market Outcomes', *Southern Economic Journal*, 80(1), pp. 26–49. Available at: <https://doi.org/10.4284/0038-4038-2011.193>.
- Fosco, G.M. and Lydon-Staley, D.M. (2020) 'Implications of Family Cohesion and Conflict for Adolescent Mood and Well-Being: Examining Within- and Between-Family Processes on a Daily Timescale', *Family Process*, 59(4), pp. 1672–1689. Available at: <https://doi.org/10.1111/famp.12515>.
- Gadermann, A.M., Guhn, M. and Zumbo, B.D. (2012) 'Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide'. Available at: <https://doi.org/10.7275/N560-J767>.
- Galassi, G., Koll, D. and Mayr, L. (2024) 'The intergenerational correlation of employment: Mothers as role models?', *Labour Economics*, 90, p. 102596. Available at: <https://doi.org/10.1016/j.labeco.2024.102596>.

- Granovetter, M.S. (1973) 'The Strength of Weak Ties', *American Journal of Sociology*, 78(6), pp. 1360–1380. Available at: <https://doi.org/10.1086/225469>.
- Heckman, J. (1974) 'Shadow Prices, Market Wages, and Labor Supply', *Econometrica*, 42(4), p. 679. Available at: <https://doi.org/10.2307/1913937>.
- Heckman, J.J. and Rubinstein, Y. (2001) 'The Importance of Noncognitive Skills: Lessons from the GED Testing Program', *American Economic Review*, 91(2), pp. 145–149. Available at: <https://doi.org/10.1257/aer.91.2.145>.
- 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), pp. 411–482. Available at: <https://doi.org/10.1086/504455>.
- Jang, K.L., Livesley, W.J. and Vernon, P.A. (1996) 'Heritability of the Big Five Personality Dimensions and Their Facets: A Twin Study', *Journal of Personality*, 64(3), pp. 577–592. Available at: <https://doi.org/10.1111/j.1467-6494.1996.tb00522.x>.
- Killingsworth, M.R. and Heckman, J.J. (1986) 'Chapter 2 Female labor supply: A survey', *Handbook of Labor Economics*. Elsevier, pp. 103–204. Available at: [https://doi.org/10.1016/S1573-4463\(86\)01005-2](https://doi.org/10.1016/S1573-4463(86)01005-2).
- Kim, J.H. and Torquati, J. (2021) 'Are You Close with Your Parents? The Mediation Effects of Parent–Child Closeness on Young Adults' Financial Socialization Through Young Adults' Self-reported Responsibility', *Journal of Family and Economic Issues*, 42(2), pp. 314–324. Available at: <https://doi.org/10.1007/s10834-020-09725-5>.
- LeBlanc, J.E. and Lyons, S.T. (2022) 'Helicopter parenting during emerging adulthood: Consequences for career identity and adaptability', *Frontiers in Psychology*, 13, p. 886979. Available at: <https://doi.org/10.3389/fpsyg.2022.886979>.
- Leppard, T.R. and Dufur, M.J. (2025) 'Family Ties and Young Adult Career Goals: Does Family Social Capital Stretch Into Emerging Adulthood?', *Emerging Adulthood*, 13(5), pp. 1219–1227. Available at: <https://doi.org/10.1177/21676968251352009>.
- Long, J.S. and Ervin, L.H. (2000) 'Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model', *The American Statistician*, 54(3), pp. 217–224. Available at: <https://doi.org/10.1080/00031305.2000.10474549>.
- MacKinnon, D.P. (2012) *Introduction to Statistical Mediation Analysis*. 1st edn. Routledge. Available at: <https://doi.org/10.4324/9780203809556>.
- Mahoney, A. (2010) 'Religion in Families, 1999–2009: A Relational Spirituality Framework', *Journal of Marriage and Family*, 72(4), pp. 805–827. Available at: <https://doi.org/10.1111/j.1741-3737.2010.00732.x>.
- Manning, A. and Petrongolo, B. (2008) 'The Part-Time Pay Penalty for Women in Britain', *The Economic Journal*, 118(526), pp. F28–F51. Available at: <https://doi.org/10.1111/j.1468-0297.2007.02115.x>.
- Millemaci, E. and Sciulli, D. (2014) 'The long-term impact of family difficulties during childhood on labor market outcomes', *Review of Economics of the Household*, 12(4), pp. 663–687. Available at: <https://doi.org/10.1007/s11150-013-9187-8>.
- Mincer, J. (1974) *Schooling, experience, and earnings*. New York: National Bureau of Economic Research; distributed by Columbia University Press (Human behavior and social institutions, 2).

- Ministry of Justice (2024) 'Official Statistics in Development: Estimates of children with a parent in prison'. Available at: <https://www.gov.uk/government/statistics/estimates-of-children-with-a-parent-in-prison/official-statistics-in-development-estimates-of-children-with-a-parent-in-prison>.
- Moffitt, R. (1984) 'The Estimation of a Joint Wage-Hours Labor Supply Model', *Journal of Labor Economics*, 2(4), pp. 550–566. Available at: <https://doi.org/10.1086/298047>.
- Next Steps (2006) 'Longitudinal Survey of Young People WAVE 3'. Available at: https://doc.ukdataservice.ac.uk/doc/5545/mrdoc/pdf/wave_three_documentation.pdf.
- Olson, D. (2011) 'FACES IV and the Circumplex Model: Validation Study', *Journal of Marital and Family Therapy*, 37(1), pp. 64–80. Available at: <https://doi.org/10.1111/j.1752-0606.2009.00175.x>.
- ONS (2026) 'Families and households in the UK'. Available at: <https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/families/bulletins/familiesandhouseholds/2025>.
- Oster, E. (2019) 'Unobservable Selection and Coefficient Stability: Theory and Evidence', *Journal of Business & Economic Statistics*, 37(2), pp. 187–204. Available at: <https://doi.org/10.1080/07350015.2016.1227711>.
- Price, J. (2008) 'Parent-Child Quality Time: Does Birth Order Matter?', *Journal of Human Resources*, 43(1), pp. 240–265. Available at: <https://doi.org/10.3368/jhr.43.1.240>.
- Roberts, B.W. and DelVecchio, W.F. (2000) 'The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies.', *Psychological Bulletin*, 126(1), pp. 3–25. Available at: <https://doi.org/10.1037/0033-2909.126.1.3>.
- Social Mobility Commission(2025) *The evolving story of social mobility in the UK: presented to Parliament pursuant to section 8B(6) of the Life Chances Act 2010*. London: Crown (State of the Nation, 2025).
- Solon, G., Haider, S. and Wooldridge, J. (2013) *What Are We Weighting For?* w18859. Cambridge, MA: National Bureau of Economic Research, p. w18859. Available at: <https://doi.org/10.3386/w18859>.
- Symonds, J.E., D'Urso, G. and Schoon, I. (2023) 'The long-term benefits of adolescent school engagement for adult educational and employment outcomes.', *Developmental Psychology*, 59(3), pp. 503–514. Available at: <https://doi.org/10.1037/dev0001458>.
- Tsoutsis, V. and Dikeos, D. (2024) 'Family Functioning and Cohesion Scale: Validation of a Short Instrument for the Assessment of Intrafamily Relations', *Behavioral Sciences*, 14(10), p. 969. Available at: <https://doi.org/10.3390/bs14100969>.
- University College London, U.I. of E. (2026) 'LSYPE1; First Longitudinal Study of Young People in EnglandNext Steps - Survey DataNext Steps: Sweeps 1-9, 2004-2023'. UK Data Service. Available at: <https://doi.org/10.5255/UKDA-SN-5545-11>.
- Wu, N. *et al.* (2018) 'Intergenerational Transmission of Educational Aspirations in Chinese Families: Identifying Mediators and Moderators', *Journal of Youth and Adolescence*, 47(6), pp. 1238–1251. Available at: <https://doi.org/10.1007/s10964-018-0820-y>.
- Zhao, L. and Zhao, W. (2022) 'Impacts of family environment on adolescents' academic achievement: The role of peer interaction quality and educational expectation gap', *Frontiers in Psychology*, 13, p. 911959. Available at: <https://doi.org/10.3389/fpsyg.2022.911959>.

9 Annex

Table 14 - Central Specification (Table 7) controls

$p < 0.1$ *, $p < 0.05$ **, $p < 0.01$ ***

Dependent variable: Ln(Wage)				
	(1)	(2)	(4)	(5)
	OLS	WLS	Heckman S2	School FE
Weekly Hours Worked (centred)	0.051*** (0.002)	0.053*** (0.002)	0.052*** (0.002)	0.051*** (0.002)
Weekly Hours Above 37.5 (FT threshold spline)	-0.038*** (0.003)	-0.041*** (0.004)	-0.040*** (0.004)	-0.041*** (0.004)
No Long-term Illness (=1)	0.068*** (0.020)	0.066*** (0.023)	0.003 (0.041)	0.056** (0.024)
Passed GCSE (=1)	0.123*** (0.021)	0.135*** (0.025)	0.099*** (0.032)	0.112*** (0.027)
Neuroticism	-0.022*** (0.004)	-0.019*** (0.004)	-0.017*** (0.004)	-0.018*** (0.004)
Conscientiousness	0.001 (0.004)	0.005 (0.005)	0.002 (0.005)	0.001 (0.005)
Extraversion	0.005 (0.004)	0.001 (0.004)	-0.002 (0.004)	0.002 (0.004)
Agreeableness	-0.002 (0.005)	-0.003 (0.006)	-0.002 (0.006)	-0.002 (0.006)
Openness	-0.004 (0.003)	0.002 (0.003)	0.002 (0.003)	0.000 (0.003)
Region: London (ref. England)	0.258*** (0.026)	0.252*** (0.032)	0.257*** (0.032)	0.230*** (0.044)
Region: Other (Wales / Scotland / N. Ireland)	0.020 (0.095)	0.016 (0.113)	0.031 (0.114)	0.033 (0.105)

Table 15 - Stage-1 Probits

 $p < 0.1$ *, $p < 0.05$ **, $p < 0.01$ ***

Dependent variable: Employed at Age 32 (=1)		
	(3)	(13)
	Heckman S1	Heckman S1
Constant	0.036 (0.281)	0.060 (0.280)
Family Cohesion Factor (FCF)	0.0830*** (0.031)	0.129** (0.053)
Male (=1)	0.256*** (0.062)	0.257*** (0.062)
Parent SES: Intermediate (ref. Low)	0.277*** (0.074)	0.255*** (0.077)
Parent SES: High (ref. Low)	0.453*** (0.072)	0.430*** (0.075)
No Long-term Illness (=1)	0.730*** (0.063)	0.729*** (0.063)
Passed GCSE (=1)	0.437*** (0.060)	0.436*** (0.060)
Neuroticism	-0.024** (0.011)	-0.024** (0.011)
Conscientiousness	0.035*** (0.012)	0.034*** (0.013)
Extraversion	0.037*** (0.012)	0.037*** (0.012)
Agreeableness	-0.014 (0.015)	-0.014 (0.015)
Openness	-0.006 (0.008)	-0.005 (0.008)
Ethnic Minority (=1)	-0.016 (0.075)	-0.019 (0.075)
Single Parent Household at Age 14 (=1)	-0.213*** (0.067)	-0.217*** (0.067)
No. of Siblings at Age 14	-0.090*** (0.026)	-0.089*** (0.026)
Region: London (ref. England)	-0.131 (0.099)	-0.131 (0.099)
Region: Other (Wales / Scotland / N. Ireland)	-0.291 (0.187)	-0.296 (0.189)
Number of Children [Exclusion]	-0.159*** (0.024)	-0.159*** (0.024)
FCF × Intermediate Parent SES	—	-0.071 (0.084)
FCF × High Parent SES	—	-0.074 (0.071)
Observations (N)	5760	5760
Wald F-test (p-value)	0.000	0.000
Pseudo R² (McFadden)	0.184	0.184

Table 16 - Parental SES Interaction Specifications

 $p < 0.1$ *, $p < 0.05$ **, $p < 0.01$ ***

Dependent variable: Ln(Wage)				
	(11)	(12)	(14)	(15)
	OLS	WLS	Heckman S2	School FE
Family Cohesion Factor (FCF)	0.019 (0.019)	0.046* (0.026)	0.032 (0.027)	0.023 (0.026)
Male (=1)	0.081*** (0.019)	0.074*** (0.023)	0.056** (0.024)	0.090*** (0.025)
Parent SES: High (ref. Low)	0.199*** (0.025)	0.153*** (0.032)	0.117*** (0.037)	0.089*** (0.032)
Parent SES: Intermediate (ref. Low)	0.081*** (0.026)	0.023 (0.033)	-0.002 (0.036)	0.040 (0.032)
FCF × High Parent SES	0.010 (0.024)	-0.027 (0.030)	-0.015 (0.031)	0.004 (0.031)
FCF × Intermediate Parent SES	0.018 (0.026)	-0.017 (0.032)	-0.007 (0.032)	0.027 (0.035)
Inverse Mills Ratio (λ)	—	—	-0.270** (0.133)	—
Observations (N)	4041	4041	4041	4005
No. of School FEs	—	—	—	608
Adj. R²	0.404	0.428	0.428	—
F-statistic	137.80	152.04	145.16	—
Adj. R² (within)	—	—	—	0.399
F-statistic (on covariates, within)	—	—	—	7.28

HC3 heteroskedasticity-robust standard-error. WLS, Heckman S2, and School FE weighted using W9FINWT.

Table 17 - P-values for diagnostic tests (Table 7)

Model	Breusch-Pagan	Ramsey RESET	Max VIF score
OLS	0.000	0.133	2.206
WLS	0.000	0.133	2.308
Heckman S2	0.000	0.270	2.206
School FE	0.000	—	2.216

Table 18 – Next Steps Wave 3 School Attitude Scale Items (w3yschat1)

Item #	Questionnaire Statement
1	I am happy when I am at school
2	School is a waste of time for me
3	School work is worth doing
4	Most of the time I don't want to go to school
5	People think my school is a good school
6	On the whole I like being at school
7	I work as hard as I can in school
8	In a lesson, I often count the minutes till it ends
9	I am bored in lessons
10	The work I do in lessons is a waste of time
11	The work I do in lessons is interesting to me
12	I get good marks for my work

*Additive score measured on a 4-point Likert scale (Strongly agree, Agree, Disagree, Strongly disagree).
Negative statements are reverse-scored such that higher values reflect more positive attitudes towards school*