Is Education the Mechanism Through Which Family Background Affects Economic Outcomes? A Generalised Approach to Mediation Analysis*

Abstract

We seek to quantify the role of education as a mechanism through which family background affects economic outcomes. To this end, we generalise mediation analysis to allow for multidimensional treatments. This improves the validity of mediation analysis for our application, in which family background is exogenous and multidimensional. Our approach allows the mediating role of education to vary across background characteristics, whilst also estimating its overall mediating effect. We estimate that educational attainment explains 21%-37% of the family background effect on hourly earnings in Australia, and only 13%-19% of the effect on wealth. We argue that these estimates are likely upward-biased. Therefore the link between family background and economic outcomes operates mostly through other mechanisms.

JEL Classification: I24; C49; J62.

Keywords: mediation analysis; intergenerational mobility; equality of opportunity

* We thank Dan Rees (the Editor), three anonymous referees, Bruce Bradbury, Paul Frijters, Craig Jones, Andrew Leigh, Chris Ryan, Nicolas Salamanca, Ian Walker and Andrew Webber, as well as seminar participants at the Australian National University, University of Melbourne, University of New South Wales, and the University of Sydney, for very useful discussions and comments on earlier versions. This work was supported by the New South Wales Department of Education.

1. Introduction

There is much consensus, internationally, on the importance of the principle of "equality of opportunity" (see Alesina and Giuliano, 2011 for a review). Whilst conceptually distinct, intergenerational mobility is perhaps the best measurable indicator we have of equality of opportunity (Corak, 2013). The extent to which children's outcomes are determined by their family background is hence a topic of considerable interest. Much progress has been made to address measurement issues and to produce internationally comparable estimates of intergenerational mobility. One common approach has been to focus on the intergenerational elasticity between male (permanent) earnings and that of their sons (Solon, 1992, Mazumder 2005, Corak, 2013, Mendolia and Siminski, 2016). An alternate approach is to focus on movements between quantiles of the earnings distribution across generations (Chetty et al., 2014).

Less progress has been made towards understanding the transmission mechanisms through which family background affects earnings, or to the policy levers which are most effective to improve mobility. A particular focus, however, is the role of education. Education is of course a major determinant of earnings (Card, 1999). And education systems are fundamentally shaped by government policy. The role of education in intergenerational mobility has been studied in three strands of the empirical literature. One strand has studied specific schooling reforms as natural experiments (Dustmann, 2004, Meghir & Palme, 2005; Holmlund, 2008; Pekkarinen *et al.*, 2009). A second strand examines the extent to which geographical variations in intergenerational mobility can be explained by differences in the characteristics of education systems (Corak 2006; 2013, Blanden 2013, Chetty *et al.*, 2014). The third, and smallest, strand of the literature has attempted to quantify the role of education as a 'mediator' – that is, the role of education as a pathway through which family background affects economic outcomes in the next generation (Bowles & Gintes, 2002, Blanden *et al.*, 2007, Kuha and Goldthorpe, 2010). Our study is in this third strand. Our main objective is to study the extent to which education is a mechanism which explains the effect of family background on earnings.

To this end, we have developed an approach which we believe to be a methodological innovation. Our innovation is a generalisation of mediation analysis (Baron and Kenny, 1986). Mediation analysis is a mainstream approach for studying causal pathways in disciplines such as statistics and psychology, and has recently been discussed formally in the

economics literature (Heckman and Pinto, 2015).[2] Standard mediation analysis seeks to estimate the extent to which the effect of a treatment (D) on an outcome (Y) is explained by a particular mechanism (M). Our innovation generalises standard mediation analysis to allow for a vector of treatment variables. We also show that standard mediation analysis is nested as a special case within our generalised approach.[3]

The motivation for our approach begins with the observation that standard indicators of intergenerational mobility are not causal parameters, which immediately makes analysis of mechanisms problematic. Consider for example the mediation analysis conducted by Blanden *et al.* (2007). Their aim was to estimate the extent to which child's education mediates the raw association between family income and child earnings. The analysis likely overstates the role of education due to positive correlations between child's education and omitted family background characteristics which also influence child earnings, and can therefore be seen as confounders. If this model is to have a causal interpretation, even the *total* effect of family income on child earnings is likely biased, due to positive correlations between family income and other omitted family background characteristics.

And yet family background is exogenous. If we were able to measure every aspect of family background perfectly (not restricting ourselves to family income), we would be able to construct unbiased estimates of the effects of family background on earnings, and then begin to explore mechanisms. Whilst it is not possible to perfectly measure family background, we can make some progress in this direction. Instead of using one indicator of family background, our proposed approach instead includes a vector of exogenous family background characteristics directly into the earnings regressions. Our approach estimates the extent to which education mediates the combined effects of all such background variables on child earnings. And it allows the mediating role of (child's) education to vary across family background characteristics. For example, child's education may have a greater role in mediating the effect of parental education than in mediating the effect of parental occupation.

---

[2] For example, see Warner (2013) and Howell (2013) for textbook treatments of mediation analysis in Statistics and Psychology, respectively.

[3] See also Tubeuf et al. (2012), who adopt a different approach to a mediation analysis with multidimensional treatment and mediator. We note however that their approach seems inconsistent with standard mediation analysis when the treatment is one-dimensional. We discuss this further in footnote 5.

We apply this approach to Australian data, exploiting the richness of the Household Income and Labour Dynamics, Australia survey data. We estimate that child's education explains around 21%-37% of the effect of family background on child (hourly) earnings. These bounds reflect different assumptions made around the relationship between education and ability. However education plays a much smaller roll in mediating the effect of family background on wealth (13%-19%). Section 2 discusses the standard mediation model and our generalisation. Section 3 discusses additional implementation issues specific to our application. Section 4 describes data and Section 5 shows results. Section 6 concludes.

2. Mediation Models

2.1 Standard Mediation Model

Identifying the mechanisms through which a given treatment (D) affects a given outcome variable (Y) is notoriously difficult. This task is known as mediation analysis in the statistics and psychology literatures, beginning with Baron and Kenny (1986). The goal of mediation analysis is to estimate the extent to which the total effect of D on Y operates through the channel of a mediating variable (M). In other words, how much of the effect of D on Y can be explained by its effect on M (the so called 'indirect' effect of D), and how much of the total effect operates through other mechanisms (the so called 'direct' effect of D).

A standard approach to mediation analysis begins by separately estimating the following two linear equations:

$$Y_i = \alpha_1 + \beta^{total} D_i + \varepsilon_i \quad (1)$$

$$Y_i = \alpha_2 + \beta^{direct} D_i + \gamma M_i + e_i \quad (2)$$

where Y_i is some outcome variable, D_i is an exogenous treatment variable and M_i is a potential mechanism through which D_i affects Y_i. β^{total} is the total effect of D_i on Y_i. β^{direct} is the component of β^{total} which does not operate through the mechanism M_i. From these estimates, the proportion of the total effect of D_i explained by mechanism M_i is given by $1 - \frac{\beta^{direct}}{\beta^{total}}$. A value of 1 suggests that mechanism M_i is the pathway through which D_i affects Y_i. A value of 0 suggests that M is not a mechanism for the effect of D_i on Y_i. It is important to note that M is not regarded as an 'omitted variable' (i.e. as a 'confounder') in

(1). Rather, D is assumed to be exogenous in (1) and estimates of β^{total} from (1) are assumed unbiased.

The important distinction between confounders and mediators is considered as fundamental in disciplines such as statistics and psychology.

Even if D_i is exogenous, mediation analysis requires strong assumptions. The key issue is that the observed mediating variable M_i is likely to be correlated with unobserved determinants of Y_i. Randomised experiments facilitate unbiased estimation of β^{total}. Nevertheless, the mediating variable is likely to be correlated with some omitted determinants of Y_i, even with experimental data. If so, the estimated effect of the mediating variable on Y_i is likely to be biased.

With observational data, the assumptions for unbiased estimation of mediation effects are of course stronger. Here, D_i is uncorrelated with other determinants of Y_i (unconfoundedness of the treatment) by assumption rather than by design. For this reason, even estimates of β^{total} are likely biased with observational data.

In our application, Y_i is the logarithm of child's earnings, D_i is some measure of family characteristics (logarithm of father's earnings being the leading candidate) and M_i is child's education. The standard mediation model may overstate the role of child's education due to positive correlations between child's education and omitted family background characteristics that also determine child's earnings. In particular, the coefficient of child's education will likely be biased upwards as it will 'pick-up' the effect of omitted family background characteristics that are positively correlated with child's education.

2.2 A Generalised Mediation Model with Multidimensional Treatment and Mediator

Our proposed extension to standard mediation analysis may improve its validity in certain circumstances. Specifically, this is where a given treatment is best thought of as multidimensional. In our application, family background incorporates a range of characteristics, including parental education and parental earnings and many other factors, which can be treated as exogenous.

For this generalised mediation approach, we replace the treatment variable and its coefficients with vectors. We also allow the mediator to be measured by a vector of indicators:

$$Y_i = \alpha_1 + \boldsymbol{\beta}^{\boldsymbol{total}}\boldsymbol{D_i} + \varepsilon_i \quad (3)$$

$$Y_i = \alpha_2 + \boldsymbol{\beta}^{\boldsymbol{direct}}\boldsymbol{D_i} + \boldsymbol{\gamma}\boldsymbol{M_i} + e_i \quad (4)$$

In our application, $\boldsymbol{D_i}$ represents all observed family background characteristics. Inevitably, $\boldsymbol{D_i}$ will not perfectly measure all relevant aspects of family background, and e_i will still contain omitted family background characteristics, which will likely be positively correlated with $\boldsymbol{D_i}$ and $\boldsymbol{M_i}$. But the extent of resulting bias will likely be smaller than in the simple mediation model in which $\boldsymbol{D_i}$ is a single variable.

Since $\boldsymbol{\beta}^{\boldsymbol{total}}$ and $\boldsymbol{\beta}^{\boldsymbol{direct}}$ are vectors, it is not trivial to summarise the overall mediating role of education. After estimating (3) and (4) separately by OLS, one can see how each element of $\widehat{\boldsymbol{\beta}}$ changes between (3) and (4), but this is not the ultimate aim. However, the overall mediating effect of $\boldsymbol{M_i}$ can be summarised by comparing the standard deviations of $\widehat{\boldsymbol{\beta}}^{\boldsymbol{total}}\boldsymbol{D_i}$ and $\widehat{\boldsymbol{\beta}}^{\boldsymbol{direct}}\boldsymbol{D_i}$, respectively, across observations in the estimation sample.[4] More specifically, the proportion of the total effect that is mediated by $\boldsymbol{M_i}$ is:

$$1 - \frac{SD(\widehat{\boldsymbol{\beta}}^{\boldsymbol{direct}}\boldsymbol{D_i})}{SD(\widehat{\boldsymbol{\beta}}^{\boldsymbol{total}}\boldsymbol{D_i})} \quad (5)$$

When the treatment is modelled as a single variable rather than a vector, this produces identical results to a comparison of $\hat{\beta}^{direct}$ and $\hat{\beta}^{total}$, as in a standard mediation model. In other words, standard mediation analysis is nested as a special case of our generalised approach.[5]

[4] This should not be confused with the standard deviations of the estimates (i.e. the standard errors).

[5] Whilst not explicit, Tubeuf et al.'s (2012) approach seems equivalent to setting $1 - \frac{Var(\widehat{\boldsymbol{\beta}}^{\boldsymbol{direct}}\boldsymbol{D_i})}{Var(\widehat{\boldsymbol{\beta}}^{\boldsymbol{total}}\boldsymbol{D_i})}$ as the mediating role of $\boldsymbol{M_i}$, when $\boldsymbol{D_i}$ is multidimensional. As we show below, using standard deviations (rather than variances) yields estimates that are consistent with standard mediation analysis when the treatment is a single variable. Mediating effects

To see this, note that β^{total} in (1) is proportional to the standard deviation of $\beta^{total}D_i$:

$$SD(\beta^{total}D_i) = \sqrt{\frac{\sum_i(\beta^{total}D_i - \overline{\beta^{total}D_i})^2}{n-1}} = \beta^{total}\sqrt{\frac{\sum_i(D_i - \overline{D_i})^2}{n-1}} \quad (6)$$

A larger effect (β^{total}) of D on Y is proportional to a higher standard deviation of $\beta^{total}D_i$, and similarly for β^{direct}. Therefore:

$$1 - \frac{SD(\beta^{direct}D_i)}{SD(\beta^{total}D_i)} = 1 - \frac{\sqrt{\frac{\sum_i(\beta^{direct}D_i - \overline{\beta^{direct}D_i})^2}{n-1}}}{\sqrt{\frac{\sum_i(\beta^{total}D_i - \overline{\beta^{total}D_i})^2}{n-1}}}$$

$$= 1 - \frac{\beta^{direct}\sqrt{\frac{\sum_i(D_i - \overline{D_i})^2}{n-1}}}{\beta^{total}\sqrt{\frac{\sum_i(D_i - \overline{D_i})^2}{n-1}}} = 1 - \frac{\beta^{direct}}{\beta^{total}} \quad (7)$$

In our application, it is useful to think of $\widehat{\boldsymbol{\beta}}^{\boldsymbol{total}}\boldsymbol{D_i}$ as an overall index of family background, as it relates to child earnings, similar to Lubotsky and Witenberg's (2006) index. An individual with a high value of $\widehat{\boldsymbol{\beta}}^{\boldsymbol{total}}\boldsymbol{D_i}$ has a family background that is associated with high expected earnings.

$\widehat{\boldsymbol{\beta}}^{\boldsymbol{total}}\boldsymbol{D_i}$ is also the predicted value from (3) and SD($\widehat{\boldsymbol{\beta}}^{\boldsymbol{total}}\boldsymbol{D_i}$) is equal to the standard deviation of these predicted values. In our application, a higher standard deviation of predicted values reflects a greater role of family background in determining child earnings. Similarly SD($\widehat{\boldsymbol{\beta}}^{\boldsymbol{direct}}\boldsymbol{D_i}$) is the standard deviation of predicted values from (4), *after* holding child education fixed. It reflects the extent to which family background determines earnings through mechanisms other than child education.

3. Implementation Issues for Estimating the Mediating Effect of Education

This section discusses a number of practical considerations for implementing this model in our application.

---

estimated using Tubeuf et al.'s (2012) approach thus cannot be interpreted in the same way as (or compared to) the results from studies which adopt the conventional mediation model.

3.1 Abilities

Ability, broadly defined, is likely to be positively correlated with socioeconomic background, child's education and child's earnings. However, there are complex causal relationships between these variables. Due to a number of factors, including home environment, parental example, genetics, *etc.* a child from well-off background may have attributes which are rewarded in the labour market, independently of educational attainment. Furthermore, high ability students tend to select into higher educational attainment. Education itself also enhances abilities. There is hence a two-way causal relationship between education and ability. It also argued that education is a mechanism which at least partially translates ability into earnings (Blanden *et al.*, 2007).

A practical consequence of this is that the approach outlined in equations (3), (4) and (5) may overestimate the mediating role of education to the extent that (i) the child's education variables are picking up the role of ability which is omitted from the model, and (ii) ability is determined by family background directly, rather than through the pathway of education. In other words, that approach may produce an upper bound for the role of education as a mediator.

An alternative approach is to control for ability in both regressions:

$$Y_i = \alpha_1 + \boldsymbol{\beta_1 D_i} + \boldsymbol{\delta_1 A_i} + \varepsilon_i \quad (8)$$

$$Y_i = \alpha_2 + \boldsymbol{\beta_2 D_i} + \boldsymbol{\delta_2 A_i} + \boldsymbol{\gamma M_i} + e_i, \quad (9)$$

Where A is a vector of ability measures.

Following same intuition as (5), the estimated share of the total effect of family background mediated by education is:

$$1 - \frac{SD(\hat{\boldsymbol{\beta}}_1 \boldsymbol{D_i}) - SD(\hat{\boldsymbol{\beta}}_2 \boldsymbol{D_i})}{SD(\hat{\boldsymbol{\beta}}^{\boldsymbol{total}} \boldsymbol{D_i})}, \quad (10)$$

where $SD(\hat{\boldsymbol{\beta}}^{\boldsymbol{total}} \boldsymbol{D_i})$ is still from (3). This approach can be regarded as a lower bound for the role of education as a mediator, since it ignores the potential role of education as a pathway through which ability is translated into earnings.

3.2 Measurement of Family Background and Child Education

As discussed above, the simple mediation model relies on strong assumptions, particularly with observational data. Our proposed generalised approach partially navigates the resulting issues by allowing more comprehensive and multidimensional measurement of family background, and by allowing the mediating role of education to vary between each of these dimensions.

Nevertheless, allowing vectors in the mediation model does not ensure that all aspects of family background will be included in the models. Indeed it may be impossible to perfectly measure all relevant aspects of family background characteristics that are relevant to child outcomes. As discussed above, such omitted family background characteristics may result in overestimation of the mediating role of education. However, it is also the case that not all aspects of educational attainment are measurable, particularly in relation to the quality of education received. This should result in underestimation of the role of education as a mediator due to classical measurement error. The net effect of these two offsetting biases is not clear.

3.3 Dimensionality Reduction

HILDA has detailed data on family background. For example, there are hundreds of parental occupation codes, hundreds of parental countries of birth and numerous variables summarising parental education. In this context, potential over-parameterisation (or 'over-fitting') is an important practical consideration. Over-fitting is the inclusion of too many parameters to be estimated in a given regression model, resulting in imprecise estimation of each parameter. This issue is typically discussed with reference to out-of-sample prediction accuracy (see for example Varian, 2014). While out-of-sample prediction is not relevant here, imprecisely estimated parameters may imply that the role of family background is not well captured in the model, despite the richness of the data. In fact, we found substantial evidence for this concern in preliminary analysis. Specifically, without dimensionality reduction, we found that the key estimates were sensitive to sample size. Smaller sample sizes (e.g. taking random sub-sets of the main estimation sample) resulted in smaller estimates for the mediating role of education.

Thus we pursued a process of reducing the number of parameters to be estimated for the large indicator variables: father's occupation; mother's occupation; father's education; mother's education; father's country of birth; and mother's country of birth. Our adopted approach is to use Lubotsky-Wittenberg indexes to summarise each of the six elements of family background (Lubotsky and Wittenberg, 2006). We discuss the construction of these indices in the appendix.

3.4 Controlling for Age and Gender

Age and gender are obviously major correlates of earnings. Whilst not shown in any of the equations above for parsimony, we also control for individual's gender and a quadratic function of age in each regression. This improves the precision of the estimates. It may also avoid bias due to potential correlations between age and family background characteristics.

4 Data

We draw primarily on the Household, Income and Labour Dynamics in Australia (HILDA) Survey, which is a representative, longitudinal study of the Australian population that started in 2001 (Wooden and Watson, 2002).

The estimation sample for the main analysis consists of 4,681 persons aged 25-54 who responded in the Wave 12 person questionnaire and who 'currently' received wages or a salary in their main job and who did not migrate to Australia after the age of five.[6] All family background variables (parents' occupation, education, country of birth, etc.) were collected as retrospective recall data from the respondent in the first wave in which they were interviewed, which for most respondents was 2001. Cognitive ability data were collected in Wave 12 for the first time. Data on non-cognitive skills were collected earlier - Big-5 personality traits data were collected in Wave 9 and locus of control data were collected in Wave 11. These were merged onto the Wave 12 data. Observations with missing values for any of the control variables were flagged with indicator variables, but retained in the estimation sample.

[6] People who migrated to Australia after the age of five were excluded from the sample because they did not conduct (all of) their schooling in Australia.

The Lubotsky-Wittenberg indexes were constructed using a larger sample of 31,625 observations across eight waves, as described in Section 5 above. Other than the larger number of waves, the same sample restrictions were applied as for the main analysis.

Key variables used in the HILDA analysis:

$\ln Y_i^{child}$ is the natural logarithm of the hourly wage of the child, derived as 'current weekly gross wages & salary in main job', divided by 'hours per week usually worked in main job'. Extreme outliers (those more than four standard deviations from the mean) were dropped. In the main estimation sample, there were 16 observations excluded on this basis, around 0.3% of the estimation sample.

$\boldsymbol{Background_i}$ is a vector of family background variables:

- Occupation of each parent (4 digit ANZSCO 2006 – which includes up to 374 categories), summarised into two Lubotsky-Wittenberg index variables (one for fathers' occupation, and one for mothers' occupation) as described in the Methods section and in the Appendix
- How much schooling each parent completed (a 5 group categorisation ranging from 'none' to 'Year 12 or equivalent') and type of post-school institution each parent received highest level qualification from (if any) (6 groups: University; Teachers College/College of Advanced Education; Institute of Technology; Technical college/TAFE; Employer; and Other), summarised into two Lubotsky-Wittenberg index variables (one each for fathers' and mothers' schooling)[7]
- Country of birth of each parent (categories for each individual country), summarised into two Lubotsky-Wittenberg index variables (one each for fathers' and mothers' country of birth)
- Aboriginal or Torres Strait Islander origin
- Age of mother at time of birth
- Whether child was living in a sole parent family at the age of 14.
- Whether father was unemployed for 6 months or more while the respondent was 'growing up'.

[7] The main results are very similar if parental years of schooling is used instead of the Lubotsky-Wittenberg parental education indices. For example, the estimated mediating roles of child's education differ by less than 1 percentage point from the preferred estimates (for each sex and overall, for upper and lower bounds).

- Number of siblings ever had

The obvious omission from the 'background' vector is parental income or earnings. Retrospective family income data were not collected in HILDA.[8] It is not clear how important this omission is. The detailed vector of other family background characteristics will be correlated with, and hence should pick up some of, the income effect. However, the omission of income suggests that the estimated importance of family background will be underestimated. The omission of family income might also lead the estimated role of education to be biased upwards, since child's education may pick up some of the family income effect that is uncorrelated with the other family background characteristics.

$\boldsymbol{Educ_i^{child}}$ is a vector of (own) educational attainment variables:

- Highest education level achieved (8 categories, ranging from Postgrad – masters of doctorate, to Year 11 and below)
- Highest year of school completed (9 categories, ranging from Year 12 to Attended primary school but did not finish, as well as a category for special needs school)
- Main field of study of highest post school qualification (15 categories, e.g. Information Technology; Law; Nursing; Creative arts)
- Which university obtained highest post school qualification from (44 categories)
- Type of school attended (government, catholic non-government, other non-government)

$\boldsymbol{Skills_i^{child}}$ is a vector of cognitive and non-cognitive skill variables:

- Three Cognitive ability variables (Backwards digits score; Word pronunciation score (short NART); Symbol-digit modalities score), as described by Wooden (2013).
- Seven Locus of Control variables, each measured on a 7-point Likert scale (e.g. 'Can do just about anything')
- Indices for each of the 'Big 5' personality traits (Agreeableness; Conscientiousness; Emotional stability; Extroversion; Openness to experience), derived from a 36 item inventory

[8] Whilst HILDA is a panel survey, it is still too short (12 years) to use direct observations of family income for people in the study population (aged 25-54 in 2012).

5 Results

5.1 The Importance of Family Background for Child Earnings

We first convey the apparent importance of family background for earnings. Table 1 summarises the distribution of predicted log hourly earnings, at various quantiles of the 'family background' distribution. The greater the dispersion of predicted values, the greater the apparent role of family background in determining earnings.

Columns (3), (4) and (5) are of primary interest. They show results corresponding with equation (3) for both genders combined and separately (after holding age constant at 40). By way of comparison, Columns (1) and (2) show additional results where (imputed) parental earnings are the only measure of family background included in the regression. In Column (1), the imputations only draw on parental occupation, similarly to Mendolia and Siminski (2016). In Column (2) the imputations are richer, drawing on each parent's occupation, education and country of birth.

Panel A of Table 1 shows various percentiles of the distribution of log earnings from each of these models. Panel B shows summary measures of these distributions. Column (1) suggests that people at the 75th percentile of parental earnings have expected earnings that are around 8 per cent higher than those at the 25th percentile. Moving from the 10th to 90th percentile of family background is associated with earnings that are 16 per cent higher. Another way of summarising this is to look at the standard deviation of predicted values, which is 0.062. Column (2), whilst using a broader set of family background characteristics in the imputation model, leads to similar conclusions.

As expected, given the richer and more flexible approach, the effect of family background is estimated to be much larger in Columns (3), (4) and (5). The model suggests that people at the 75th percentile of 'family background' have expected earnings that are 21.5% higher than those at the 25th percentile. People at the 90th percentile of 'family background' have expected earnings that are 39% higher than those at the 10th percentile. When each gender is

analysed separately, family background matters even more (for both sexes).[9] Males at the 90th percentile have expected earnings that are 56.5% higher than those at the 10th percentile. For females, the corresponding difference is also large (52.6%). The standard deviations of predicted log earnings from the preferred model are 0.160 overall, 0.188 for males and 0.181 for females. These are more than twice as large as those in columns (1) and (2).

These results presented in Table 1 are of substantive interest. Their main implication is that models which draw only on parental earnings (or at least imputed earnings) greatly understate the importance of family background for child earnings. To the extent that child's education is correlated with those unmeasured family background factors, a standard mediation analysis will consequently overestimate the mediating role of education. These results lend support for using a multidimensional measure of family background.

5.2 The Role of Education as a Mediator

Notes: This table summarises the importance of family background as a determinant of earnings. It shows the distribution of predicted values from regressions of ln(earnings) on family background indicators, holding age and sex constant. In columns (1) and (2) family background is measured by imputed parental earnings. In column (1), this imputation draws only on each parent's occupation. In column (2) the imputation draws on each parent's occupation, education and country of birth. Columns (3), (4) and (5) show results from regression models which include a vector of family background characteristics, including parental occupation, education, country of birth and other variables, described in full in Section **Error! Reference source not found.**. Panel A shows percentiles of the distributions of predicted earnings. Panel B shows summary statistics on those distributions. Greater dispersion of predicted values is indicative of a larger estimated role of family background in determining earnings.

[9] This is despite the fact that gender is controlled for in the analysis when both sexes are combined. A likely explanation is that various aspects of family background matter differently for males and for females and so the specification in the combined-gender analysis is too restrictive.

Table 2 shows the key results, which summarise the importance of education as a mediator of family background's effect on earnings. For each model, the table shows an 'upper bound' (estimated using a model which ignores cognitive and non-cognitive skills) and a 'lower bound' (estimated using a model which ignores the role of education as a pathway for skills to influence earnings). As discussed above, the lower bound is $1 - \frac{SD(\hat{\boldsymbol{\beta}}_1 \boldsymbol{D_i}) - SD(\hat{\boldsymbol{\beta}}_2 \boldsymbol{D_i})}{SD(\hat{\boldsymbol{\beta}}^{\boldsymbol{total}} \boldsymbol{D_i})}$ and the upper bound is $1 - \frac{SD(\hat{\boldsymbol{\beta}}^{\boldsymbol{direct}} \boldsymbol{D_i})}{SD(\hat{\boldsymbol{\beta}}^{\boldsymbol{total}} \boldsymbol{D_i})}$.

Columns (1) and (2) show results where (imputed) parental earnings are the only measures of family background included in the models for comparative purposes. In Column (1), the imputations only draw on parental occupation, whilst in Column (2) the imputations draw on each parent's occupation, education and country of birth. The other columns show results for the preferred model which directly includes all family background characteristics.

As expected, the estimated role of education is largest in Column (1), followed by Columns (2) and then Column (3). Column (1) implies that education accounts for between 33% and 66% of intergenerational transmission and does not differ greatly by gender. In Column (2), the mediating role of education is slightly smaller (between 31% and 62%) and is considerably larger for females than for males.

Results from the preferred model are in Column (3). They suggest that education accounts for between 21% and 37% of the family background effect on earnings. This suggests that education has a substantial role in explaining intergenerational transmission. However, the majority of the family background effect is transmitted through other mechanisms. These 'other mechanisms' may include intergenerational transmission of personal attributes (either through genetics or through environment), including (cognitive and non-cognitive) skills, as well as transmission of preferences over work versus leisure. Access to social capital networks may also contribute. The results also suggest the mediating role of education may be slightly greater for females than for males. Whilst family background is a stronger determinant of earnings for males than for females (Table 1), its effect on educational attainment is more similar for each sex (as will be shown in Section 5.3).

The remainder of Table 2 considers the mediating role of education for some of the key dimensions of family background, still drawing on the results from the preferred model. For these results, we are simply comparing pairs of individual parameters, before and after controlling for child's education. This is similar to the 'standard' approach to mediation

analysis, except that other family background characteristics are controlled for in each regression. As hypothesised, the mediating role of own education is largest for the effect of parental education and this is especially the case for females. In the 'upper bound' results with both sexes combined, own education is estimated to mediate 74% of the effect of father's education and 60% for mother's education. For females, it is even larger (77%) for the effect of mother's education.

Appendix B includes a comparative analysis of Australia and Britain of the mediating role of education. It shows that education may play a larger mediating role in Australian than in Britain.

In Table 3 we further explore the mediating role of education, by considering several additional economic outcome measures: annual earnings, annual personal income, annual household income and household net worth. We also consider whether the results are sensitive to the using three-year averages of each outcome variable, rather than the single-wave measure we have used in other results.[1] Across these outcome variables, the largest estimated mediating effects of education are for hourly earnings. (Table 3, Columns (1) and (6)). The mediating effect is somewhat smaller for annual earnings, annual personal income, and annual household income (which are all quite similar), and smaller again for household net worth. This seems sensible, since education is a major and direct determinant of human capital (and hence the hourly wage rate), but its effect on the other outcome variables is less direct. Human capital is only one determinant of annual earnings and of total income. Therefore it is sensible for mediating role of education to be smaller for annual earnings and total income than for hourly earnings. Finally, income is only one determinant of wealth, and so it seems sensible that the mediating role of education to be smaller again for household net worth.

[1] For each outcome variable except net worth, Table 3 shows results for outcomes measured at wave 12 and for a three-year average between waves 10 and 12. For net worth (which is measured only every 4 years), results are shown for wave 14, and for the average of net worth at waves 10 and 14. Extreme outliers (those more than 4 standard deviations away from the mean of the outcome variable being investigated) are excluded in each case, affecting around 0.5% of the sample in each case.

5.3 Does the Education System Promote Intergenerational Persistence or Mobility?

The main analysis above suggests that education 'explains' some component of the effect of family background on earnings. This positions education as 'part of the problem' rather than 'part of the solution' to intergenerational transmission of advantage. In a sense, this is a correct interpretation to the extent that people from disadvantaged backgrounds receive less schooling. However, it is informative to consider the *extent* to which family background determines educational outcomes, and compare this to the extent to which family background determines earnings. In other words, we know that family background is a major determinant of earnings, but is family background a *smaller* determinant of educational attainment? If so, then perhaps one can gauge the extent to which the education system is actually facilitating intergenerational mobility rather than contributing to intergenerational persistence.

To this end, we repeated the analysis that underlies Table 1, this time with educational attainment (instead of earnings) as the dependent variable.[2] The first measure we used is ln(years of schooling). This is a simple and transparent summary measure of educational attainment. The limitation of this measure, however, is that it ignores many aspects of educational attainment which may be related to both earnings and to family background. This includes school sector (private; catholic; public), as well as field and institution of tertiary education. Thus we created a second dependent variable, which is an educational attainment index. This variable summarises all available educational attainment variables into a single Lubotsky-Wittenberg index, using weights which correspond to the estimated relationship between each educational variable and own earnings.[3]

The key results from both versions are shown in Table 4. This table shows summary statistics for the dispersion of predicted educational attainment, similar to what was shown for earnings in the lower panel of Table 1. The upper panel of Table 4 shows results for educational attainment measured in logarithm of years of schooling. It suggests that people at the higher end of the background distribution are expected to receive considerably more education. For example, those at the 90th family background percentile can expect to receive 27% more

[2] To mirror the main analysis, the Lubotsky-Wittenberg family background indexes were re-created using all 8 waves of data, with ln(years of education) used as the dependent variable.

[3] Specifically, this is the predicted value from a regression of ln(hourly earnings) on all available educational attainment variables (as detailed in the data section which describes key HILDA variables), after controlling for sex and a quadratic in age. Only Wave 12 was used as it has all of the required variables.

years of schooling (approximately 3 more years) compared to those at the 10th percentile.[4] The corresponding discrepancy is slightly larger for females than for males.

The more comprehensive education index is used in the lower panel. Here the importance of background is larger still (as expected). Those on the 90th background percentile can expect to receive 36% more schooling. Interestingly, the difference between genders is small here, and if anything the importance of background is larger for men. While family background has a larger effect on the quantity of schooling for women (upper panel), this is offset by the types of education induced. This presumably relates to field and institution of tertiary study, perhaps also in terms of secondary school sector.

These results should be compared to the corresponding (Model 3) results in Table 1. This comparison reveals that family background is a considerably smaller determinant of educational attainment than the corresponding relationship between family background and earnings. Comparing the P90 – P10 results for both genders combined, the family background effect is around 30% smaller for educational attainment than the family background effect for earnings.[5] A comparison of P75 – P25 results leads to a similar conclusion (31%). Comparisons of the other summary measures also give similar results. To reiterate, family background has a smaller role in determining educational attainment than it does in determining earnings. In this sense, the education system is 'part of the solution' rather than 'part of the problem' in intergenerational transmission of economic advantage.

6 Conclusion

We have taken a 'big picture' view on the role of education in intergenerational economic mobility. This is a topic of immense policy interest, which also comes with enormous methodological challenges. It is clearly impossible to randomly assign family background. Even if it were possible, to estimate the role of education as a transmission mechanism would still require major assumptions. In this context, our observational analysis should be seen as a modest attempt to make a piecemeal contribution to a very complex topic. In doing so, we have developed a new methodological approach for summarising the extent to which

[4] If we use years of schooling in levels (instead of in logs) as the dependent variable, we get very similar results. For example, this version of the model suggests that those at the 90th family background percentile can expect to receive 25% more years of schooling than those at the 10th percentile.

[5] (1 – 36%/51.6%) × 100%.

education mediates the effect of all observed family background characteristics on earnings. Our innovation is a generalisation of mediation analysis. In our generalisation, the treatment and mediator can be multidimensional constructs. We have argued that this approach improves the validity of mediation analysis in this application, even if it does not completely eradicate all potential sources of bias.

Our results suggest that family background is a major determinant of economic wellbeing in Australia. Further, there is a positive relationship between family background and education, and a positive relationship between education and earnings. It follows that education is one of the mechanisms through which economic advantage is transferred from one generation to the next.

The main results suggest that education may explain around 21%-37% of the effect that family background has on hourly earnings. The upper bound (37%) is estimated using models which ignore cognitive and non-cognitive skills (which are correlated with both education and family background). Conversely, the lower bound (21%) is estimated using models which ignore the role of education as a pathway through which traits influence earnings.. However, economic advantage is transmitted between generations mainly through other mechanisms. Further, the mediating role of education is smaller for other measures of economic outcomes, particularly for wealth (between 13% and 19%). This reflects the fact that human capital (which education creates directly) is only one determinant of wealth. Overall, the role of education as a mechanism is not large.

Perhaps the greatest remaining threat to the validity of the analysis is potential correlation between child education and other mediators of the family background effect. This should bias the estimated mediating role of education upward. There is also likely bias due to measurement error in family background, ability, and education. Our results suggest that the direction of bias due to measurement error is positive for family background and ability, and negative for education. Of these three constructs, education is the most tangible and probably best measured. So we think that the sources of positive bias are very likely to dominate the source of negative bias. And our analysis generally finds that the mediating role of education is relatively small nonetheless, and much smaller than implied by simpler methods. Therefore we conclude that the family-background effect on earnings is mostly due to mechanisms other than education.

We also attempted a comparable analysis for the United Kingdom. The results suggest that family background determines earnings to a similar degree in the two countries. They also suggest that the mediating role of education may be larger for Australia. These results should be interpreted cautiously, since there are considerable differences in the data sources which could not be avoided. Further research is required for confident conclusions on differences in the mediating effect of education between countries.

Appendix A: Dimensionality Reduction

This appendix describes the process by which rich data on parental occupation, parental country of birth and parental education were summarised into six indices, in order to avoid problems of over-parameterisation discussed in Section 3.3.

There are numerous approaches to dimensionality reduction. The simplest approach here would be to use higher levels of aggregation for each classification. For example, to use a 3-digit occupational classification rather than the more detailed 4-digit classification. In general, higher levels of aggregation result in a larger estimated role of education in explaining the family background effect. A concern with such an approach, however, is the loss of detail in measuring family background. The unmeasured component of family background may be correlated with child's education. Thus the role of education may be over-estimated for the same reasons that we raised in relation to the simple mediation model. Principal component analysis (factor analysis) was also considered, but this is not a useful technique when the dimensionality issue is characterised by mutually exclusive dummy variables, which are by construction uncorrelated with each other.

Our preferred approach is to use Lubotsky-Wittenberg indexes to summarise each of the 6 family background characteristics listed above (Lubotsky and Wittenberg, 2006). Each index is a weighted sum of the original indicators. The weights applied were the parameter estimates from an un-reduced version of the regression model represented by equation (3). In other words, the weight applied to each indicator variable is proportional to the strength of that indicator's association with child earnings. These indexes were then used in place of the indicator variables for all of the regression models. Instead of estimating 607 parameters in

the domains of parental occupation, education and country of birth, we are left with just six parameters in these domains after dimensionality reduction.

This approach implicitly invokes a restriction on the original specification – for a given indicator variable (e.g. occupation), the effect of each category is assumed to change proportionally between equations. In other words, the mediating role of child's education in the effect of fathers' occupation is assumed to be constant across occupational categories, and similarly for the other indicator variables. This restriction comes at a cost – it does not allow for meaningful heterogeneity-analysis between sections of the background distribution. For example, we cannot confidently address the important question of whether education plays a greater role for intergenerational transmission at the top vs the bottom of the family background distribution. However, we believe that this approach yields more credible estimates of the overall mediating effect of education.

In preliminary analysis, we conducted this reduction technique 'in-sample'. But this did not eliminate the sample-size sensitivity. In the preferred analysis, we instead constructed these indexes using the parameter estimates (as weights) from a regression with the largest possible appropriate sample. This sample consists of the eight waves of HILDA that have the required data to estimate equation (3). Thus we used eight times more data to construct more precise weights for the index construction. This amounts to having better (less noisy) measures of family background in the analysis. This approach yields results which are not sensitive to the sample size used in the main regressions. This seems to be the most effective way to address the dimensionality issue whilst retaining the richness of the available data on family background.

Appendix B: Comparative Analysis between Australia and Great Britain

We compare the results from HILDA with corresponding results derived from the British Cohort Study (BCS). BCS is a survey of more than 17,000 children born in Great Britain between 4th and 11th April 1970. The survey has followed the lives of these individuals and collected information on health, physical, educational and social development and economic circumstances of their families. Since the birth surveys, there have been seven waves of data, with information collected at age 5, 10, 16, 26, 30, 34 and 42. Employees were asked to provide information on their usual pay, pay period, and hours usually worked in a week. We use this information to derive hourly earnings at age 26, 30, 34 and 42.

We also use data on individual educational qualifications and we construct a vector $\boldsymbol{Educ_i^{child}}$, including the information on the highest qualification attained at every wave (6 groups, ranging from Post-degree qualification to Low High School graduate). Various parental background characteristics were collected at every wave and we use information on parental age and marital status at birth, country of birth and parental occupation and education when the child was 16.

In the analysis performed with BCS, $\boldsymbol{Background_i}$ is a vector of family background variables including:

- Occupation of each parent (which includes around 300 categories)[6]
- How much schooling each parent completed (a 7 group categorisation ranging from 'none' to Degree or equivalent)
- Region of birth for each parent (12 categories representing countries or groups of countries)
- Age of mother at time of birth
- Whether child was living in a sole parent family at birth.

[6] Detailed data on parental occupation were collected through the Family Follow Up Form in 1986. This form was not completed by 20% of the sample, who were excluded from the analysis. The reasons for failure to complete the form are not known, raising concerns over potential sample selection bias. This may reduce comparability of results between HILDA and BCS. An earlier version of this analysis did not exclude those observations, instead flagging them with an indicator variable (Mendolia and Siminski, 2015).

We construct a panel data set, by pooling all the different waves of BCS data and using data on individual earnings at age 26, 30, 34, 38 and 42. The estimation sample consists of 17,180 observations. At each wave, employees are asked to report their usual pay, the pay period, and the hours usually worked in a week. We use this information to construct hourly earnings. Observations for individuals who are self-employed are dropped from the analysis. Parental education and occupation are derived from information collected when the child was 16. The model also includes information on both parents' region of birth, marital status and age of the mother when the child was born. At each wave, information on the child's highest academic qualification is also collected. Standard errors in all regressions are clustered on the individual to account for multiple observations per individual used in each model.

Following Blanden *et al.* (2007) we perform factor analysis on several variables collecting behavioural ratings. We then include in the model a vector $\boldsymbol{Skills_i^{child}}$ of cognitive and non-cognitive skill variables including:

- antisocial and neurotic behaviour at age 5
- English Picture Vocabulary test (EPVT) and a copying test administered at age 5
- Indicators of behaviours at age 10:
    - antisocial attitude
    - clumsiness
    - concentration
    - extroversion
    - hyperactivity
    - anxiety
- A reading and a maths test administered at age 10.

HILDA (Comparable-with-BCS version)

We also estimate a second version of the HILDA analysis which is intended to be as comparable as possible to the BCS analysis. This involves limiting the sample to the set of persons aged 26-42 and excluding persons born overseas. These sample restrictions leave 2,550 observations for the main analysis and 17,240 observations for the L-W index creation.

This version also involves collapsing some of the explanatory variables or dropping variables from the Background vectors and especially the Education vector. The modified versions of these are shown below:

Comparable-to-BCS $\boldsymbol{Background_i}$ variables:

- Occupation of each parent (4 digit ANZSCO 2006) summarised into two Lubotsky-Wittenberg index variables (one for fathers' occupation, and one for mothers' occupation) as described in the Methods section, above.
- How much schooling each parent completed (a 3 group categorisation: Year 10 or below; Year 11 or equivalent; Year 12 or equivalent) and type of post-school institution each parent received highest level qualification from (if any) (6 groups: University; Teachers College/College of Advanced Education; Institute of Technology; Technical college/TAFE; Employer; and Other), summarised into two Lubotsky-Wittenberg index variables (one each for fathers' and mothers' schooling)
- Country of birth of each parent (collapsed into 10 categories), summarised into two Lubotsky-Wittenberg index variables (one each for fathers' and mothers' country of birth)
- Age of mother at time of birth
- Whether child was living in a sole parent family at the age of 14.

Comparable-to-BCS $\boldsymbol{Educ_i^{child}}$ variables:

- Highest education level achieved (5 categories, ranging from Postgrad – masters of doctorate, to Year 11 and below)

The cognitive and non-cognitive skills vector was unchanged despite major comparability issues, explicitly because we sought to judge whether the inclusion of HILDA's measures have similar effects on the results as compared to that of the superior skills measures in the BCS.

Results

We first compare the importance of family background as a determinant of earnings in the two countries. We then consider the role of education as a mediator of the family background

effect for the two countries. Finally, we also seek to gain insights into whether the (inferior) set of cognitive and non-cognitive traits in HILDA are serving their intended purpose. That is, we are interested in whether the inclusion of traits measured at early childhood (as are included in BCS) impacts the results differently to the inclusion of traits measured contemporaneously with wages (as are included in HILDA).

The estimated importance of parental background on child earnings in both countries is summarised in Table A.1.[7] The results suggest that family background has a similar role in explaining child earnings for the two countries. For example, the BCS analysis suggests that people at the 90th (75th) percentile of 'family background' have expected earnings that are around 53% (24%) higher than those at the 10th (25th) percentile. The corresponding estimate is 54% (26%) in HILDA. The standard deviations of these predicted earnings distributions are 0.18 for both countries. This contrasts with work that suggests family background (proxied by fathers' earnings alone) has a greater effect on child earnings in the UK than it does in Australia (Corak 2013; Mendolia and Siminski, 2016).

Table A.2 shows the percentage of the family background effect that is explained by child's education for both countries, similarly to the main analysis shown in Table 2. It suggests that the mediating effect of education may be larger in Australia. The results suggest that education accounts for between 14% and 26% of the family background effect in Australia, compared to between 11% and 22% in the UK. The mediating effects are also larger for Australia when each gender is analysed separately. The estimated role of education for Australia is smaller in Table than in the main results (Table 2). This is to be expected because the main analysis includes a much richer set of own-education variables.

Further, cognitive and non-cognitive skills do not have a systematically larger role in explaining intergenerational transmission in BCS as compared to HILDA. This can be seen by comparing the difference between the lower bound and upper bound estimates of the role of education in BCS and in HILDA in Table . For example, this difference equals 11 percentage points in the combined gender analysis in BCS, and 12 percentage points in HILDA. This is despite the much higher quality data on traits collected in BCS. There is

[7] An earlier version of this analysis (Mendolia and Siminski, 2015) contained an error for BCS. That version summarised the distribution of predicted earnings after controlling for child education, thereby understating the 'total' effect of family background.

hence no evidence that the lower quality traits measures in HILDA result in biased lower bounds of the Australian results.

References

Alesina A, Giuliano P, Bisin A and Benhabib J (2011) 'Preferences for Redistribution' in Benhabib J, Bisin A and Jackson M (eds) *Handbook of Social Economics*, North Holland, 93-132.

Baron R and Kenny D (1986) 'The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations' *Journal of Personality and Social Psychology* 51, 1173–1182.

Blanden J, Gregg P and Macmillan L (2007) 'Accounting for intergenerational income persistence: noncognitive skills, ability and education' *Economic Journal*, 117, C43-C60.

Blanden J (2013) 'Cross-country ranking in intergenerational mobility: a comparison of approaches from economics and sociology' *Journal of Economic Surveys*, 27, 38-73.

Bowles S and Gintis H (2002) 'The Inheritance of Inequality' *The Journal of Economic Perspectives*, 16 (3): 3-30.

Card D (1999) 'The Causal Effect of Education on Earnings', in Ashenfelter, O and Card, D (eds) *Handbook of Labor Economics* Vol 3A, pp 1801-1863.

Chetty R, Hendren N, Kline P and Saez E (2014) 'Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States' *Quarterly Journal of Economics*, 129 (4): 1553-1623.

Corak M (2006) 'Do poor children become poor adults? Lessons from a Cross Country Comparison of Generational Earnings Mobility' *Research on Economic Inequality* 13, 143-188.

Corak M (2013) 'Income Inequality, Equality of Opportunity, and Intergenerational Mobility' *The Journal of Economic Perspectives*, 27 (3), 79-102.

Dustmann C (2004) ‘Parental background, secondary school track choice, and wages’ *Oxford Economic Papers*, 56, 209–230.

Heckman J and Pinto R (2015) ‘Econometric Mediation Analyses: Identifying the Sources of Treatment Effects from Experimentally Estimated Production Technologies with Unmeasured and Mismeasured Inputs’ *Econometric Reviews* 34 (1-2): 6-31.

Holmlund H (2008) *Intergenerational Mobility and Assortative Mating: Effects of an Educational Reform*, Centre for Economics of Education, London School of Economics, CEE DP 91.

Howell (2013) *Statistical Methods For Psychology*, 8th ed., Cengage.

Kuha, J. and Goldthorpe, J.H. (2010) Path analysis for discrete variables: the role of education in social mobility. *Journal of Royal Statistical Society Series A* 173: 351-369.

Lubotsky D and Wittenberg M (2006) ‘Interpretation of Regressions with Multiple Proxies’ *The Review of Economics and Statistics* 88: 549-562.

Mazumder B (2005) ‘The Apple Falls Even Closer to the Tree than We Thought: New and Revised Estimates of the Intergenerational Inheritance of Earnings’ in Bowles S Gintis H and Osborne Groves M (eds.) *Unequal Chances: Family Background and Economic Success*, New York: Russell Sage Foundation.

Meghir C and Palme M (2005) ‘Educational reform, ability, and parental background’ *American Economic Review* 95, 414–424.

Mendolia S and Siminski P (2015) *The Role of Education in Intergenerational Economic Mobility in Australia*, Revised Final Report, commissioned by the NSW Government Office of Education – Centre for Education Statistics and Evaluation (CESE).

Mendolia S and Siminski P (2016) ‘New Estimates of Intergenerational Mobility in Australia’ *Economic Record*, 92, 361–373.

Pekkarinen T, Uusitalo R and Kerr S (2009) ‘School tracking and intergenerational income mobility: Evidence from the Finnish comprehensive school reform’ *Journal of Public Economic*s 93, 965-973.

Solon G (1992) ‘Intergenerational Income Mobility in the United States’ *American Economic Review* 82 (3): 393-408.

Tubeuf S, Jusot F, Bricard D (2012) ‘Mediating Role of Education and Lifestyles in the Relationship between Early-Life Conditions and Health: Evidence from the 1958 British Cohort’ *Health Economics* 21 (Suppl. 1): 129-150.

Varian H (2014) ‘Big Data: New Tricks for Econometrics’ *Journal of Economic Perspectives,* 28 (2): 3-28.

Warner R (2013) *Applied Statistics: From Bivariate Through Multivariate Techniques*, 2/E, Sage.

Table 1 – The Importance of Family Background as a Determinant of Earnings

	Using Parents' Earnings only V1	Using Parents' Earnings only V2	Preferred Model – Using all Family Background Characteristics		
	Both genders	Both genders	Both genders	Males	Females
	(1)	(2)	(3)	(4)	(5)
A: Percentiles of Predicted Log Earnings Distribution					
p1	3.207	3.198	3.056	3.057	2.957
p5	3.320	3.313	3.180	3.210	3.070
p10	3.342	3.344	3.231	3.286	3.132
p15	3.366	3.369	3.269	3.332	3.170
p20	3.380	3.383	3.298	3.364	3.199
p25	3.390	3.397	3.321	3.393	3.223
p30	3.402	3.409	3.341	3.415	3.241
p35	3.411	3.418	3.358	3.437	3.264
p40	3.418	3.425	3.374	3.455	3.283
p45	3.424	3.433	3.391	3.474	3.307
p50	3.433	3.440	3.407	3.495	3.328
p55	3.439	3.448	3.427	3.517	3.347
p60	3.446	3.455	3.446	3.534	3.369
p65	3.451	3.462	3.467	3.560	3.394
p70	3.458	3.469	3.490	3.586	3.424
p75	3.467	3.476	3.515	3.611	3.452
p80	3.474	3.485	3.543	3.639	3.483
p85	3.483	3.492	3.577	3.679	3.516
p90	3.491	3.504	3.625	3.733	3.554
p95	3.505	3.523	3.687	3.808	3.633
p99	3.548	3.569	3.874	4.017	3.819
B: Summary Measures of Predicted Log Earnings Dispersion					
P60 - P40	0.028	0.029	0.072	0.078	0.086
expressed as % difference in expected hourly earnings	2.8%	3.0%	7.4%	8.2%	8.9%
P75 - P25	0.077	0.079	0.195	0.217	0.229
expressed as % difference in expected hourly earnings	8.0%	8.2%	21.5%	24.3%	25.7%
P90 - P10	0.149	0.160	0.394	0.448	0.422
expressed as % difference in expected hourly earnings	16.1%	17.3%	48.3%	56.5%	52.6%
Standard deviation of predicted log hourly earnings distribution	0.062	0.068	0.160	0.188	0.181

Notes: This table summarises the importance of family background as a determinant of earnings. It shows the distribution of predicted values from regressions of ln(earnings) on family background indicators, holding age and sex constant. In columns (1) and (2) family background is measured by imputed parental earnings. In column (1), this imputation draws only on each parent's occupation. In column (2) the imputation draws on each parent's occupation, education and country of birth. Columns (3), (4) and (5) show results from regression models which include a vector of family background characteristics, including parental occupation, education, country of birth and other variables, described in full in Section **Error! Reference source not found.**. Panel A shows percentiles of the distributions of predicted earnings. Panel B shows summary statistics on those distributions. Greater dispersion of predicted values is indicative of a larger estimated role of family background in determining earnings.

Table 2 – The Role of Education as a Mediator of the Relationship between Family Background and Hourly Earnings

	Using Parental Earnings only		Preferred Model – Using all Family Background Characteristics						
Estimated Mediating Role of Education	V1 (1)	V2 (2)	Overall (3)	Father's Education (4)	Mother's Education (5)	Father's Occupation (6)	Mother's Occupation (7)	Father's Country of Birth (8)	Mother's Country of Birth (9)
				A: Both genders (of child)					
Lower bound	33%	31%	21%	40%	34%	17%	18%	27%	22%
Upper bound	66%	62%	37%	74%	60%	29%	29%	37%	33%
				B: Males					
Lower bound	29%	22%	14%	20%	20%	12%	12%	23%	14%
Upper bound	63%	53%	28%	44%	40%	24%	20%	37%	24%
				C: Females					
Lower bound	32%	39%	25%	46%	50%	23%	20%	26%	29%
Upper bound	66%	71%	35%	70%	77%	31%	27%	33%	35%

Notes: This table shows the estimated importance of child's education as a mechanism through which family background affects hourly earnings. A value of 100% implies that education is the sole mechanism through which family background affects earnings, while 0% suggests that education is not a mechanism for the family background effect on earnings. Columns (1) and (2) show results from models in which family background is measured only by imputed parental earnings. In column (1), this imputation draws only on each parent's occupation. In column (2) the imputation draws on each parent's occupation, education and country of birth. Columns (3)-(9) show results from regression models which include a vector of family background characteristics, including parental occupation, education, country of birth and other variables,

described in full in Section **Error! Reference source not found.**. Column (3) shows the main results, summarising the role of education in mediating the overall effect of family background. Columns (4)-(9) show results from the same models as (3). They show the extent to which education mediates the effects of key elements of family background. The 'upper bounds' are estimated using models which ignore cognitive and non-cognitive skills (which are correlated with both education and family background). The 'lower bounds' are estimated using models which ignore the role of education as a pathway through which cognitive and non-cognitive skills influence earnings.

Table 3 – The Role of Education as a Mediator of the Relationship between Family Background and Various Outcome Measures

Estimated Mediating Role of Education	Single-year outcome measures					Three-year average outcome measures				
	Hourly Earnings (1)	Annual Earnings (2)	Personal Annual Income (3)	Household Annual Income (4)	Household Net Worth* (5)	Hourly Earnings (6)	Annual Earnings (7)	Personal Annual Income (8)	Household Annual Income (9)	Household Net Worth* (10)
			Both genders (of child)					Both genders (of child)		
lower bound	21%	19%	17%	19%	13%	20%	17%	16%	18%	11%
upper bound	37%	31%	30%	30%	19%	35%	29%	29%	29%	18%
			Males					Males		
lower bound	14%	16%	12%	17%	12%	14%	13%	12%	16%	12%
upper bound	28%	25%	21%	25%	15%	26%	21%	22%	24%	15%
			Females					Females		
lower bound	25%	12%	11%	14%	10%	22%	14%	10%	13%	9%
upper bound	35%	21%	20%	22%	15%	31%	23%	18%	21%	15%

Notes: This table shows the estimated importance of child's education as a mechanism through which family background affects various outcome measures. A value of 100% implies that education is the sole mechanism, while 0% suggests that education is not a mechanism for the family background effect. Column (1) shows the same results as Table 2, Column (3). The results in the other columns are also from models which use the complete set of family background characteristics, but with different outcome variables. Columns (1) - (5) us single-wave measures for each outcome variable: Wave 12 for columns (1) – (4); and Wave 14 for column (5) because wealth was only measured in every

fourth wave. Columns (6) - (9) use 3-year averages of these same outcomes (Waves 10-12), while column (10) uses the average across waves 10 and 14. See also Table 2 notes.

Table 4 – The Importance of Family Background for Educational Attainment

	Both genders (1)	Males (2)	Females (3)
A: Dependent Variable: ln(years of schooling)			
P60 - P40	0.043	0.043	0.051
expressed as % difference in expected years of schooling	4.4%	4.3%	5.3%
P75 - P25	0.117	0.116	0.136
expressed as % difference in expected years of schooling	12.4%	12.3%	14.6%
P90 - P10	0.239	0.256	0.272
expressed as % difference in expected years of schooling	27.0%	29.2%	31.3%
Standard Deviation of predicted log years of schooling	0.095	0.101	0.108
B: Dependent Variable: L-W Education Index			
P60 - P40	0.053	0.058	0.057
expressed as % difference in expected educational attainment	5.4%	5.9%	5.9%
P75 - P25	0.151	0.152	0.156
expressed as % difference in expected educational attainment	16.2%	16.4%	16.8%
P90 - P10	0.307	0.334	0.308
expressed as % difference in expected educational attainment	36.0%	39.7%	36.0%
Standard Deviation of predicted Education Index	0.121	0.132	0.121

Notes: This table summarises the importance of family background as a determinant of educational attainment. It shows summary statistics on the distribution of predicted values from regressions of educational attainment on a vector of family background characteristics, including parental occupation, education, country of birth and other variables, described in full in Section **Error! Reference source not found.**, holding age and sex constant. For Panel A, the dependent variable is the natural logarithm of years of schooling. For Panel B, the dependent variable is an educational attainment index, which draws on all available data on quality and quantity of education. It is constructed through a Lubotsky-Wittenberg procedure using weights which correspond to the estimated relationship between each educational variable and earnings. In both panels, a greater dispersion of predicted values is indicative of a larger estimated role of family background in determining educational attainment.

Table A.1 –Family Background as a Determinant of Earnings – Comparison of UK and Australia

	Both genders (1)	Males (2)	Females (3)
A: United Kingdom (BCS)			
P60 - P40	0.085	0.100	0.085
expressed as % difference in expected wage	8.9%	10.5%	8.9%
P75 - P25	0.218	0.269	0.233
expressed as % difference in expected wage	24.4%	30.9%	26.3%
P90 - P10	0.424	0.516	0.464
expressed as % difference in expected wage	52.9%	67.6%	59.1%
Standard deviation of predicted log hourly wage distribution	0.180	0.228	0.203
B: Australia ('Comparable' HILDA)			
P60 - P40	0.081	0.083	0.099
expressed as % difference in expected wage	8.5%	8.7%	10.4%
P75 - P25	0.228	0.239	0.251
expressed as % difference in expected wage	25.6%	27.0%	28.5%
P90 - P10	0.431	0.479	0.472
expressed as % difference in expected wage	53.9%	61.4%	60.3%
Standard deviation of predicted log hourly wage distribution	0.182	0.232	0.204

Notes: This table summarises the importance of family background as a determinant of earnings in the United Kingdom and Australia. The results for Australia differ from those in Table 1, due to a number restrictions made here to the sample and the variable set. These changes were made to improve comparability with the BCS data. The table shows summary statistics on the distribution of predicted values from regressions of earnings on a vector of family background characteristics, including parental occupation, education, country of birth and other variables, described in full in Sections **Error! Reference source not found.** and **Error! Reference source not found.**, holding age and sex constant. A greater dispersion of predicted values is indicative of a larger estimated role of family background in determining earnings.

Table A.2 – The Mediating Role of Education – Comparison of UK and Australia

	UK (BCS)	Australia (HILDA, 'comparable' analysis)
	(1)	(2)
	A: Both genders (of child)	
Lower bound	11%	14%
Upper bound	22%	26%
	B: Males	
Lower bound	5%	7%
Upper bound	13%	16%
	C: Females	
Lower bound	11%	16%
Upper bound	20%	24%

Notes: This table shows the estimated importance of child's education as a mechanism through which family background affects earnings in the United Kingdom and Australia. The results for Australia differ from those in Table 2, due to a number restrictions made here to the sample and the variable set. These changes were made to improve comparability with the BCS data. A value of 100% implies that education is the sole mechanism through which family background effects earnings, while 0% suggests that education is not a mechanism for the family background effect on earnings. The results are from regression models which include a vector of family background characteristics, including parental occupation, education, country of birth and other variables, described in full in Sections **Error! Reference source not found.** and **Error! Reference source not found.**. The 'upper bounds' are estimated using models which ignore cognitive and non-cognitive skills (which are correlated with both education and family background). The 'lower bounds' are estimated using models which ignore the role of education as a pathway through which skills influence earnings.