J Econ Inequal (2018) 16:103–122
https://doi.org/10.1007/s10888-017-9358-6

Bridging the gaps: inequalities in children's educational outcomes in Ireland

David Madden[1]

Received: 21 November 2014 / Accepted: 4 September 2017 / Published online: 30 September 2017

Abstract Recent developments in the inequality literature have stressed the importance of inequality of opportunity as opposed to inequality of outcome. In this paper we investigate the presence of inequality of opportunity in two measures of educational achievement for a representative sample of Irish 9 year olds. Students are partitioned into four groups according to maternal education levels and gaps in outcomes are calculated between each group. Quantile decompositions of the pairwise gaps reveal substantial gaps between groups and that almost half of the gaps can be explained by differences in characteristics between the groups. Detailed decompositions show consistently significant effects for income, number of childrens books in the home and maternal age.

Keywords Quantile decomposition · Inequality in education

1 Introduction

Education surely plays a pivotal role in many key outcomes in life, such as earnings, career choice and health (see Ashenfelter et al. 1999, for evidence on the relationship between education and earnings, Cutler and Lleras-Muney 2010, with respect to health and Oreopolous and Salvanes 2011, for other non-pecuniary returns). Education can also provide substantial positive externalities to society in general (see the review and references in Dickson and Harmon 2011). Given these benefits of education to both the individual and society, it seems important that all individuals have the opportunity to acquire education. A corollary of this

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10888-017-9358-6) contains supplementary material, which is available to authorized users.

David Madden
david.madden@ucd.ie

1 School of Economics, University College Dublin, Belfield, Dublin 4, Ireland

is that arbitrary circumstances of background should not act as an impediment to the acquisition of education and that educational opportunities and achievements should not differ according to arbitrary circumstances. A further corollary of this position is that if two people from different backgrounds have access to the same level of educational resources, their opportunity to translate these resources into educational attainment should be the same. What each individual ultimately makes of the educational opportunities presented to them in terms of their innate abilities and the effort they expend may be regarded as a private concern, but from the point of view of society, it seems desirable at the least that all citizens should have the opportunity to invest in their education and that the return to a given investment of effort should not differ by arbitrary circumstance.

Yet there is ample evidence that educational opportunities and achievements differ systematically according to family socioeconomic status (SES) (for a comprehensive survey, see Björklund and Salvanes 2010) and hence that such equality of opportunity does not exist. Note that given the age at which formal education takes place (typically starting at age 5 and ending at 18 or perhaps 21 if someone goes on to third level) it seems to reasonable to assume that family SES are, to a large extent, out of the control of the student. Much of this research has focussed on participation in third level education (see for example Denny 2010, for Ireland). However, there is also evidence that the socioeconomic gradient may set in at much earlier ages (Feinstein 2003; Cunha and Heckman 2007).

The contribution of this paper is to examine such a socioeconomic gradient in test scores in mathematics and reading for a nationally representative sample of nine year olds in Ireland. However, there are two novel aspects to our approach. As our measure of SES, we use the educational level of the mother. Using this measure, we partition our sample into four ordered, mutually exclusive and exhaustive categories and a detailed Blinder-Oaxaca (BO) type decomposition is applied to the pairwise gaps in test scores between each of the categories. This enables us to look behind the pairwise gaps to investigate the role of characteristics (or endowments) and returns to characteristics.

Secondly, rather than the standard BO decomposition which is evaluated at the mean, we use the recentered influence function (RIF) regression approach of Firpo et al. (2009) to carry out the detailed decomposition at different quantiles of the distribution. Decompositions evaluated at the mean can miss important information as to what is happening elsewhere in the distribution, in particular in parts of the distribution which may be of more concern to policy-makers.

The remainder of the paper is as follows. In the next section we provide some more background on differences in educational outcomes by SES. We also outline the BO decomposition approach, particularly when applied at different quantiles of the distribution. We also discuss possible interpretations of this richer analysis. In Section 3 we describe our data and in Section 4 we present our results, while Section 5 provides concluding comments and discussion.

2 Inequality of opportunity in education and the Blinder-Oaxaca decomposition

There is abundant evidence that ex post, equality of achievement by SES is generally not found. In their comprehensive review of the relationship between education and family background, Oreopolous and Salvanes (2011) list two motivations for concern over this relationship. The first is what they term the *equality of opportunity* motivation. If there is to be equality of opportunity in education, then arbitrary circumstances over which a person has no control, such as family SES, should not be a relevant factor in determining educational outcomes. The second motivation is what they term the *child development* motivation, whereby family SES influences parental resources and thus how much parents can invest in their children's education.

Our analysis in this paper will shed light on both of these issues. We show the presence of inequality of opportunity via the gaps in educational achievement by family SES in Ireland, gaps which appear at a very young age. However, the detailed BO decomposition also permits a deeper analysis of the possible factors lying behind these gaps. While it must be stressed that detailed BO decompositions do *not* permit a causal interpretation, they do point towards factors which are statistically associated with gaps in educational achievement and hence areas which may be usefully explored for future policy initiatives. In particular, it is possible to examine the relative role played by household and school factors.

Furthermore, the role played by *returns* to characteristics may also be critical. Thus, while we will see that higher SES families who have "higher" or "better" characteristics (in terms of more abundant endowments) have better educational outcomes, it may also be the case that higher SES families can obtain a greater return from any given set of characteristics. This may reflect a role for unobserved characteristics and/or greater efficiency in the use of observed characteristics by higher SES families. Such a phenomenon is consistent with the notion of complementarities between capabilities and investment as outlined by Cunha and Heckman (2007) whereby early investments in a child will influence her capacity to learn and may manifest itself in higher educational returns to a given set of school and home characteristics. In turn this may demand a different policy response, as simply providing extra resources to schools or lower SES families may not be sufficient to bridge the gap.

Our analysis of these gaps using quantile decomposition will also provide extra insight and may prove important from a policy perspective. It seems reasonable that an inequality-averse policy-maker who wishes to bridge the educational gaps between children with different SES will have greater concerns for those children with low educational achievements. Those children will have very low absolute levels of educational achievement and possibly very poor lifetime prospects.

Having provided a motivation for our approach in the form of detailed decomposition of gaps between families of SES, we now explain how we carry out this analysis. The BO approach partitions the population into mutually exclusive and exhaustive types, on the basis of some observed characteristic and a reduced form equation for the outcome under consideration is estimated for each type. Many applications of the BO approach examine outcomes in labour markets and so the reduced form equation is typically a wage equation, but in our application it can be regarded as a reduced form education production function. Since linear regressions hold exactly at the mean, BO showed that for a comparison between any two types, the gap in outcomes at the mean could be decomposed exactly into that part arising from characteristics and that part arising from the return to characteristics. The former is often referred to as the "explained gap", while the latter is the "unexplained gap" and can also, depending upon the particular application, be viewed as a measure of a treatment effect (Fortin et al. 2011).

More formally, Blinder (1973) and Oaxaca (1973) show that, given an outcome, y_t (e.g. a test score) for students in type t (t=1,2), and assuming

$$y_t = X_t'\beta_t + v_{t,}, E(v_t) = 0, t \in \{1, 2\}$$

then a decomposition of the difference in mean outcomes between the two types is provided by

$$\begin{aligned}\Delta_y^{\mu} &= E(y_1) - E(y_2) = E(X_1)'\beta_1 - E(X_1)'\beta_2 + E(X_1)'\beta_2 - E(X_2)'\beta_2, \\ &= (E(X_1)'[\beta_1 - \beta_2]) + ([E(X_1)' - E(X_2)']\beta_2).\end{aligned}$$

where $E(X_1)\beta_2$ is the unconditional counterfactual mean outcome i.e. what type 1 would have achieved on average if they had the returns of type 2. X represents a vector of characteristics and β is a vector of returns to characteristics (or slope parameters of the relationship, including the intercept).

The second term on the right hand side above, $([E(X_1) - E(X_2)]\beta_2)$, shows that part of the gap which arises owing to differences in the characteristics of the two groups and is sometimes referred to as the "explained" portion of the gap. The first term on the right hand side,$(E(X_1)[\beta_1 - \beta_2])$, is that part of the gap which arises owing to differences in the returns to characteristics, and is sometimes referred to as the "unexplained" portion of the gap. It is also possible to further decompose both the explained and unexplained portions of the gap to obtain the contribution of each covariate. This is sometimes called the "detailed decomposition".[1,2]

Note that in the decomposition above, in the explained portion of the gap, the differences in characteristics are weighted by the returns from group 2. An alternative decomposition, essentially the mirror image of the decomposition above, is also possible where the difference in characteristics are this time weighted by the returns from group 1. The key issue here is essentially the choice of a reference vector of returns coefficients which can be regarded in some sense as neutral or non-discriminatory between the two groups.

We choose the higher achieving group as our reference vector as it seems more likely that lower achieving groups are being "discriminated" against, or alternatively that policy-makers in wishing to move towards greater equality would favour a levelling up, rather than a levelling down.[3]

However, we may also be interested in gaps and decompositions at parts of the distribution other than the mean. Unfortunately, the simple BO decomposition holds exactly only at the mean, and so we need an alternative approach in order to carry out regression-based decompositions in the spirit of BO at different quantiles.

A number of approaches to this issue have been proposed (see the review by Fortin et al. 2011). Given our outcome, y, the conditional quantile function is assumed to be linear of the form

$$Q_\theta(y\,|X) = X_i'\beta_\theta \text{ for each } \theta \in (0, 1),$$

[1]Detailed decompositions of the unexplained portion can also be sensitive to the choice of omitted category for categorical variables. See Fortin et al. (2011).

[2]It is also possible to have a three-way decomposition. This recognises the fact that typically both characteristics and returns will differ between the two groups simultaneously, and so a third interaction term takes account of this. The inclusion of a third interaction term in the context of a decomposition which is already analysing a sub-component of inequality of opportunity (that part between differences in types) seems to present an extra layer of confusion, especially as interpretation of this term can be difficult. Thus it was decided to proceed with a two-way decomposition. I am grateful to an anonymous referee for raising this point.

[3]Thus in the analysis below we use the returns for the higher achieving group as the reference. However we also investigated the sensitivity of the results to the choice of a pooled reference vector i.e. the vector of coefficients of returns is that obtained from a pooled regression over both groups. In all cases the qualitative results were very similar and these results are available on request. I am grateful to an anonymous referee for raising this point.

where X_i represents the set of covariates for individual i and β_θ is the coefficient vector for the θ^{th} quantile. The quantile coefficients can be seen as capturing the return of each covariate across the distribution of y. Given the assumption of linearity, it is possible to estimate the conditional quantile of y by linear quantile regression for each $\theta \in (0, 1)$. The conditional quantiles for types 1 and 2 are then $Q_{1\theta}(y_1 | X_1) = X'_{1i}\beta_{1\theta}$ and $Q_{2\theta}(y_2 | X_2) = X'_{2i}\beta_{2\theta}$ respectively.

However, in order to carry out the BO decomposition at different quantiles, it is necessary to obtain the quantiles of the unconditional distribution, which can then be used to construct the counterfactual which is fundamental to the decomposition. This is straightforward when dealing with the mean, since the law of iterated expectations tells us that $E(y) = E_X(E(y | X))$ and hence the OLS estimate for covariate X_i provides the effect of the covariate on either the conditional or unconditional mean of y. However, critically the law of iterated expectations does not hold in the case of quantiles and so $Q_\theta(y) \neq E_X[Q_\theta(y | X)]$ where $Q_\theta(y)$ is the θ^{th} quantile of the unconditional distribution and $E_X[Q_\theta(y | X)]$ is the corresponding conditional quantile. Thus, in terms of a decomposition, the differences in unconditional quantiles will not be the same as the difference in conditional quantiles and hence it is not straightforward to recover (and decompose) the gap between unconditional quantiles. A number of approaches have been suggested to overcome this and we choose to follow that of Firpo et al. (henceforth FFL, 2009) which additionally is capable of providing a detailed decomposition.

FFL suggest an OLS-based regression method which estimates the impact of changes in an explanatory variable on the unconditional quantile of the outcome variable, via the regression of a transformation of the outcome variable on the set of explanatory variables. The transformation in question is based on the *influence function* (IF), which provides the influence of an individual observation on the distributional statistic of interest (such as the variance, or a particular quantile). In the case of the mean, for example, the influence function is the demeaned value of the outcome variable i.e. $y - \mu$. What is known as the re-centered influence function (RIF) is obtained if the original distributional statistic of interest is added back to the IF. Thus in the case of the mean, the RIF $= y - \mu + \mu = y$.

More generally (and dropping type subscripts for convenience), if $F(y)$ is the cumulative distribution of the outcome variable and if $T(.)$ is the distributional statistic in question, e.g., a quantile, then the influence function is the directional derivative of $T(F)$ at F (Essama-Nssah and Lambert 2011). By adding the IF to the original distributional statistic, we obtain the RIF. By construction, the RIF obeys the law of iterated expectations and thus $E[RIF(y; T(.), F(y)] = T(.)$ and it is this which is regressed against the covariates in the X vector.

For the case where the distributional statistic is a specific quantile, Q_θ, the IF is defined as

$$IF(y; Q_\theta) = \frac{(\theta - I[(y \leq Q_\theta])}{f_y(Q_\theta)},$$

where θ is the quantile in question, $I(.)$ is an indicator function taking on the value of 1 if the expression in parentheses is satisfied, $Q_\theta(y)$ is the θ^{th} quantile of the unconditional distribution of the outcome variable and $f_y Q_\theta(y)$ is the density of the marginal distribution of y evaluated at Q_θ (see Essama-Nssah and Lambert 2011). The RIF is then

$$RIF(y; Q_\theta) = Q_\theta + \frac{(\theta - I[y \leq Q_\theta])}{f_y(Q_\theta)}.$$

A brief numerical example may be helpful here. Suppose we wish to calculate the value of the RIF for a test score, and let the distributional statistic in question be the median, i.e., $\theta =$

0.5. For illustration we will calculate the value of the RIF for three observations, those at the 25th, 50th and 75th percentile and assume their values, and the values of density function are as in the table below (this example is taken from the excellent survey by Porter (2015).

y	Quantile	θ	Q_θ	$I(y \leq Q_\theta)$	$f_y(Q_\theta)$	RIF
45.5	.25	.50	54	1	0.0353	39.855
54	.50	.50	54	1	0.0353	39.855
60	.75	.50	54	0	0.0353	68.144

For the first observation of y, at the 25th percentile, the value of 45.5 is clearly less than or equal to Q_θ and so the indicator function takes on a value of 1, and hence the second term in the expression for the RIF is negative. This implies that the value of the RIF for this observation (39.855) is less than the original value of the observation. However for the value of y at the 75th percentile, clearly $y \geq Q_\theta$ and hence the indicator function is zero. In this instance the second term in the RIF is positive and hence the value of the RIF (68.144) for this observation is greater than the original value of the observation. Note that in this example, we have assumed that the density f_y takes on the same value for all y. In practice this will need to be estimated using kernel density methods.

Having calculated the value of RIF for all observations in this way, the RIF regression model is then defined as $E[RIF(y;\, Q_\theta)\,|X] = X'\beta$, and can be estimated by OLS. The estimated coefficients of the vector β then give the effect of each covariate on the unconditional θ^{th} quantile of y. The regression can be estimated for different values of θ and for different types, and counterfactuals can be constructed as with the standard BO decomposition, including a detailed decomposition (though the omitted category issue remains).[4]

Before concluding this section, we discuss some possible interpretations of the approach taken in this paper. In the literature on inequality of opportunity, a distinction is made between what may be regarded as "fair" and "unfair" sources of inequality (see for example Romer 2013; Rosa Dias and Jones 2007). For example, what are sometimes labelled as "circumstances" such as parental socio-economic circumstances are seen as unfair sources of inequality, whereas inequality arising from factors such as effort or lifestyles may be seen as fair.

In order to detect the presence of ex post inequality of opportunity between different types, Romer (2013) proposes to measure the gaps in outcomes between types, evaluated at the same quantile. Lying behind this approach is what is known as the Roemer Identification Assumption, whereby individuals of different types but at the same quantile within their type are viewed as expending the same *degree* (italics in original) of effort. Thus, the absolute level of effort may differ, but individuals in different types but ranked at the same within type quantile are viewed as expending the same degree of effort. There are clear parallels with our analysis here where we decompose the gaps at specific quantiles into explained and unexplained portions and in certain applications quantile decomposition may be an attractive approach to exploring the Roemer view of ex post inequality of opportunity. There is however one key problem in applying this interpretation to our specific example here, and that is the concept of "effort" as it applies to children.

Is it reasonable to assume that nine year old children (as in the application here) consciously exert effort and that they should, in some sense, be held responsible for their effort? The literature on inequality of opportunity has not always been consistent in its treatment

[4] For a recent application of the FFL approach which compares it to other decomposition approaches, see Baltagi and Ghosh (2015).

Table 1 Principal carers' education

Education level	Principal carer (%)
1. Primary/lower secondary	29.4
2. Complete secondary	37.3
3. Post school, non-degree	16.2
4. Primary degree	17.1
Total	100

of this issue (as pointed out by Kanbur and Wagstaff 2015) but Roemer has stated that he believes that it is not appropriate to regard anyone under the age of 16 being held responsible for their effort.

Nevertheless, accepting that it is not appropriate to regard children as exerting effort in the sense that it may be applied to adults, it still may be useful to carry out the decomposition of the gaps in outcomes and make the distinction between characteristics and unexplained factors. The contribution of this paper, which measures and further decomposes such gaps at different quantiles, is still of interest as it seems highly likely that policy-makers might wish to distinguish between inequality of opportunity at "high" and "low" levels of outcome, and the quantile decompositions carried out here enable us to do this, as well as calculating the contribution of individual factors.[5]

Finally, in some applications of the BO approach, the unexplained portion of the gap is referred to as "discrimination". In the context of this paper, it could also be viewed as an extra dimension of inequality suffered by children from low SES families and indeed seems to be close to inequality of opportunity as discussed in Breen and Goldthorpe (1999, 2001). They maintain that "...children of disadvantaged class origins have to display *far more merit* than do children of more advantaged origins in order to attain similar class positions (Breen and Goldthorpe 1999)". They interpret merit as a combination of ability and effort, and one interpretation of their result is that having the same characteristics and expending the same level of effort is not sufficient to bridge gaps in outcomes between children of different backgrounds. Notwithstanding the difficulty of applying the concept of effort to children, this interpretation could be applied to the unexplained portion of the gaps evaluated at given quantiles.

3 Data and summary statistics

Our data come from the Growing Up in Ireland (GUI) Survey 9 year old cohort which tracked the development of a cohort of children born in Ireland in the period November 1997–October 1998 (see Williams et al. 2009). The sampling frame of the data was the national primary school system, with 910 randomly selected schools participating in the study. Part of the survey consisted of the children undertaking tests in mathematics and reading which were administered by the GUI fieldworkers at the school. These tests are known in Ireland as the Drumcondra tests and have been a feature of the Irish educational system for a number of years and are linked to the national curriculum. These are administered on

[5] A similar point is implicitly made by Kanbur and Wagstaff (2015) when discussing differences in outcomes at levels of destitution.

Table 2 Summary drumcondra logit scores by gender (standard deviation in brackets)

	Total	Female	Male
Maths	– 0.759	– 0.822	– 0.699
Reading	0.012 (0.994)	0.015 (0.965)	0.009 (1.020)

an annual basis to all children in the primary school system. However, the particular tests for the GUI survey had not been seen by schools, teachers or pupils in advance of their use in GUI, thus it seems unlikely that students would have been intensively prepared for these tests, although they would have had some familiarity with tests of this kind from previous years.[6] In addition, the Drumcondra tests have no implications for further progression in the school system. The particular cohort of nine year olds in the GUI survey were spread over three different school grades (2nd, 3rd and 4th class) and three different levels of the test were administered, with the majority of the children in 3rd class (roughly equivalent to grade 3 in the US).

The educational outcome which we use in this paper are the results from these tests in maths and reading. As the tests were administered at three different levels, it was necessary to standardise the results, hence the data we use are the logit scores which were obtained from the original raw data using the principles of Item Response Theory (see Lord 1980). Results from tests at this age (and earlier) have been shown to have predictive power for subsequent later-life outcomes in areas such as education and health (Feinstein 2003).

In total there are 8568 children in the GUI survey. As our definition of "type", we use the education level of the principal carer. We drop observations where the Drumcondra test results were missing (222 observations). We also drop observations where the principal carer is not the biological mother of the child (210 observations). In carrying out our decompositions, we employ a wide range of variables which might influence the test scores. These include data on the study child's principal carer, family and school circumstances. Where these are missing, we drop those observations (see the appendix for a detailed list of variables employed). This gives us a sample for analysis of 7536 of which 3663 are boys and 3873 are girls. In all cases sampling weights are applied.[7]

We construct our types on the basis of mother's education and we divide this into four categories. Category 1 is those who have completed no further than lower secondary school education, indicating that they left formal schooling at or before the age of 16. Category 2 is those who completed secondary schooling, thus leaving formal education at around age 18. Category 3 is those who have taken a post-school, but non-degree, qualification, while category 4 is those with at least a primary degree. While a finer breakdown by education was available, we chose to limit ourselves to four types, as a higher number of types would have reduced cell size and would also have added to the number of pairwise decompositions. Table 1 summarises educational qualifications for mothers.

[6]For more details on these tests see Murray et al. (2011).

[7]The variable with the greatest number of missing observations was family income. To address this we replaced these missing observations via conditional mean imputation. The inclusion/non-inclusion of these observations made little qualitative difference to the results.

Table 3 Mean scores by type (SD in brackets)

	Maths	Reading
1. Primary/ low sec	− 1.121	− 0.355
2. Complete secondary	− 0.707	0.009 (0.948)
3. Non-degree	− 0.616	0.169 (0.970)
4. Primary Degree	− 0.385	0.500 (0.907)

Table 2 provides the average logit scores for maths and reading by gender. Girls show higher average scores for reading, while boys show higher average scores for maths. For

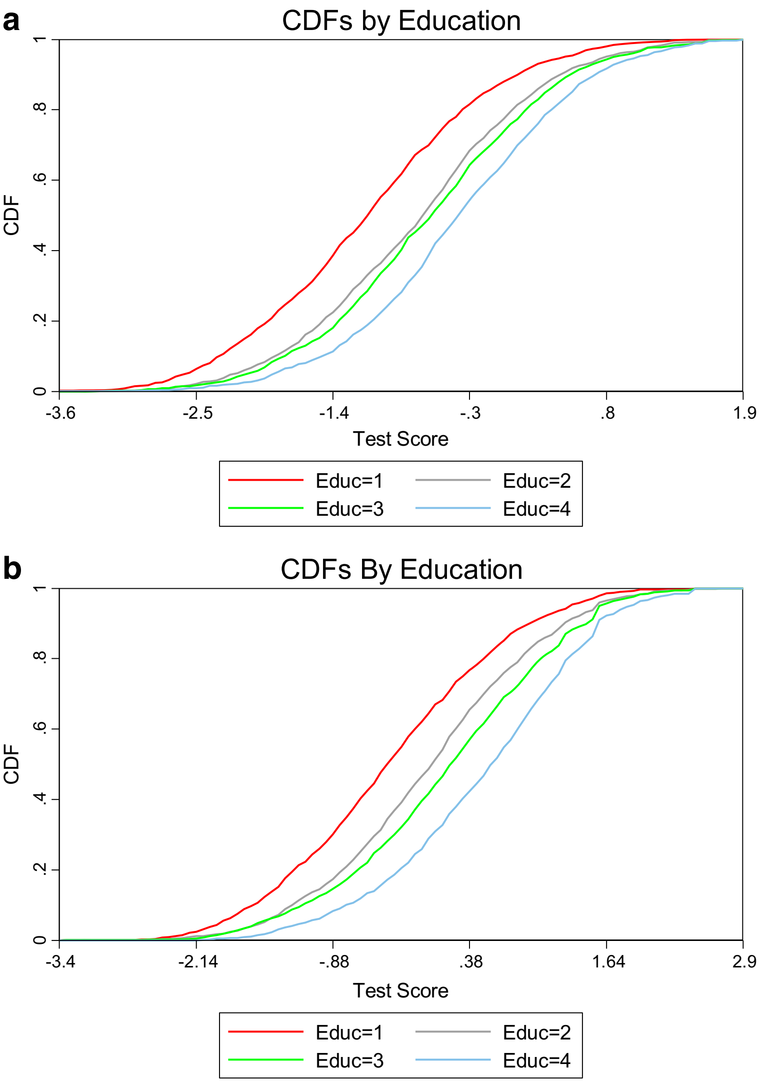

Fig. 1 **a** CDFs maths scores by maternal education type. **b** CDFs reading scores by maternal education type

Table 4 Definition of variables

Variable	Definition
Age	Age of principal carer of study child
Birthweight	Study child's birthweight in kg
Early birth	0/1 variable, takes value of 1 if study child born at 32 weeks or earlier.
Smoker	0/1 variable, takes value of 1 if principal carer is current smoker
Preg smoker	0/1 variable, takes value of 1 if principal carer was daily smoker during pregnancy
Preg drinker	0/1 variable, takes value of 1 if principal carer drank weekly or more during pregnancy
Breastfed	0/1 variable, takes value of 1 if study child was ever breastfed
Illness	0/1 variable, takes value of 1 if study child has ongoing chronic illness
Log Eqinc	Log of Equivalised Household Income
Mum healthy	0/1 variable, takes on 1 if self-assessed health of principal carer is excellent, very good or good.
Trauma	Sum of answers to 0/1 questions relating to whether study child experienced a range of traumas including death of parent/close relative, divorce/separation of parents, serious injury of family member, drug-taking/alcoholism in immediate family etc
Books	Categorical (1–5) response to question of number of childrens books which study child has access to in home
Local 1	Sum of answers to categorical (1–5) questions regarding quality of local area in terms of litter, vandalism, drug-taking etc
Local 2	Sum of answers to categorical (1–5) questions regarding how safe for children to play in area etc
Working	0/1 variable relating to whether or not principal carer is working or not
Size class	Total number of children in study child's class, numeric ranging from 13 to 36
Par/teacher	0/1 variable relating to whether parent attends parent-teacher meeting
Engage	Variable reflecting teachers engagement with class in terms of monitoring progress, variable is the sum of 5 0/1 questions, with weekly monitoring taking value of 1, less frequent monitoring taking value of 0
Texperience	Numeric variable, number of years teacher has been teaching at primary level
quality	Numeric variable reflecting quality of school facilities (based on response of principal) – school principal asked 17 questions regarding school quality. Variable is sum of "excellent" responses, ranging from 0 to 17
School size	Ordinal numeric variable (1–10) reflecting size of school, ranging from 1–80 pupils to >400
Young Sibling	Number of younger siblings
Old Sibling	Number of Older Siblings
Partner	0/1 variable reflecting whether principal carer has partner in household

Table 5 Summary statistics – mean and standard deviation (in parenthesis)

Variable	Full Sample	1. Prim/ Low Sec	2. Sec. Educ	3. Non Degree	4. Degree
Age	39.016	38.110	39.147	38.699	40.591
	(5.533)	*(5.827)*	*(5.416)*	*(5.424)*	*(4.986)*
Birthweight	3.512	3.437	3.537	3.559	3.543
	(0.621)	*(0.631)*	*(0.616)*	*(0.580)*	*(0.641)*
Early birth	0.017	0.024	0.014	0.015	0.011
	(0.128)	*(0.153)*	*(0.119)*	*(0.120)*	*(0.104)*
Smoker	0.250	0.414	0.208	0.189	0.119
	(0.433)	*(0.492)*	*(0.405)*	*(0.391)*	*(0.324)*
Smoked while pregnant	0.163	0.310	0.129	0.090	0.054
	(0.369)	*(0.462)*	*(0.335)*	*(0.286)*	*(0.226)*
Drank while pregnant	0.014	0.014	0.014	0.013	0.014
	(0.117)	*(0.117)*	*(0.118)*	*(0.115)*	*(0.117)*
Breastfed	0.446	0.240	0.420	0.572	0.735
	(0.497)	*(0.427)*	*(0.493)*	*(0.494)*	*(0.441)*
Illness	0.105	0.134	0.093	0.101	0.083
	(0.306)	*(0.340)*	*(0.290)*	*(0.300)*	*(0.275)*
Equivalised Income (log)	9.702	9.437	9.713	9.818	10.024
	(0.518)	*(0.504)*	*(0.446)*	*(0.439)*	*(0.526)*
Mother Healthy	0.933	0.882	0.951	0.955	0.962
	(0.249)	*(0.322)*	*(0.215)*	*(0.207)*	*(0.191)*
Trauma	1.493	1.506	1.412	1.591	1.555
	(1.300)	*(1.339)*	*(1.267)*	*(1.316)*	*(1.275)*
Books in house	4.171	3.775	4.208	4.379	4.570
	(1.080)	*(1.217)*	*(1.036)*	*(0.926)*	*(0.810)*
Local 1	−12.658	−11.957	−12.897	−12.807	−13.198
	(2.764)	*(3.063)*	*(2.684)*	*(2.625)*	*(2.247)*
Local 2	6.413	6.545	6.399	6.390	6.237
	(1.842)	*(1.765)*	*(1.906)*	*(1.803)*	*(1.854)*
Mother Working	0.540	0.397	0.546	0.596	0.718
	(0.498)	*(0.489)*	*(0.498)*	*(0.490)*	*(0.449)*
Class size	26.031	24.848	26.354	26.540	26.873
	(6.392)	*(7.148)*	*(6.101)*	*(6.124)*	*(5.558)*
Parent/teacher meeting	0.885	0.852	0.903	0.886	0.897
	(0.319)	*(0.354)*	*(0.295)*	*(0.317)*	*(0.304)*
Engagement	2.861	2.854	2.856	2.869	2.876
	(0.870)	*(0.879)*	*(0.884)*	*(0.837)*	*(0.854)*
Teacher Experience	12.753	12.47041	13.238	12.420	12.497
	(11.293)	*(11.473)*	*(11.397)*	*(11.056)*	*(10.947)*
Teacher Qualifications	2.416	2.211	2.398	2.401	2.822
	(2.963)	*(2.741)*	*(2.995)*	*(2.971)*	*(3.208)*
School size	402.333	429.419	329.386	541.978	382.788
	(1951.098)	*(2014.927)*	*(1769.535)*	*(2252.441)*	*(1903.889)*

Table 5 (continued)

Variable	Full Sample	1. Prim/ Low Sec	2. Sec. Educ	3. Non Degree	4. Degree
Younger sibling	0.782	0.723	0.739	0.843	0.922
	(0.864)	*(0.867)*	*(0.832)*	*(0.858)*	*(0.913)*
Older sibling	0.958	1.124	0.942	0.815	0.846
	(0.989)	*(1.059)*	*(0.954)*	*(0.929)*	*(0.956)*
Partner in household	0.835	0.783	0.853	0.836	0.882
	(0.371)	*(0.412)*	*(0.353)*	*(0.370)*	*0.322*

subsequent analysis, we do not differentiate by gender (for analysis of the differential achievement by gender for maths, see Doris et al. 2013).

In Table 3 we present the mean results for maths and reading scores by type. In all cases average scores in maths and reading are higher for those children whose mothers' have higher levels of education. At the extremes, the gap between the most advantaged and least advantaged types approaches one standard deviation of score. In terms of comparison, it should be noted that such gaps are larger than the gaps observed between ethnic groups in the US (e.g. the Black-White or Hispanic-White gaps) for similar tests for similar age groups (see, for example, Clotfelter et al. 2009, who analyse gaps between grades 3 and 8 in the US). The importance of such gaps in cognitive/educational outcomes in terms of future adult outcomes has been explored by Hanushek (1986) and Haveman and Wolfe (1993). Low achievement in childhood tends to persist and significantly worse life outcomes as adults may result.

Figure 1a and b show cumulative distribution functions (cdfs) for the test scores. The cdf for education type 4 is well to the right and below those of the other types. That for type 1 is well to the left and above, while the cdfs for types 2 and 3 are quite close together. The horizontal gap between the cdfs for each education type reflects the gap in scores at that quantile and so these graphs indicate a reasonable degree of ex post inequality of opportunity between type 1 and the other types, and also between type 4 and the other types. In Section 4, we propose to investigate these gaps in more detail via quantile decomposition.

Finally, Table 4 describes the characteristics used in the BO decomposition, while Table 5 provides summary statistics for the complete sample and by education level of the mother. We note that children whose mothers had the lowest educational level tend to have lower birthweight and were less likely to have been breastfed. Their mothers were also more likely to be smokers and to have poor health. In addition, these children were growing up in poorer households, households with less books and with a lower frequency of a partner present. Mothers for this group were also less likely to be working outside the home although class sizes were smaller. The gradient for most variables within the other three types is less pronounced.

GUI is a rich dataset and we are able to include a wide variety of observed factors. However, inevitably there will also be unobserved factors which will influence the outcome, and since we are unable to observe them, their impact will come under the "unexplained" heading. Thus, what we list below as the explained portion of the gap should be regarded as a lower bound, since presumably if unobserved factors became observable their impact would be reflected in the explained portion (I am grateful to an anonymous referee for this point).

Table 6 Quantile decompositions, maths

Quantile	Total test score gap	Explained (percentage)	Unexplained (percentage)	Income (percentage)	Books (percentage)
Types 4 and 3 (primary degree versus non-degree)					
10	0.316	39	62	3	6
25	0.264	47	53	18	7
50	0.224	59	41	20	13
75	0.219	34	67	16	9
90	0.210	36	64	10	12
Mean	0.231	49	51	11	11
Types 4 and 2 (primary degree versus complete secondary)					
10	0.386	51	49	5	9
25	0.358	51	49	21	9
50	0.308	65	35	23	19
75	0.324	31	69	17	11
90	0.308	27	73	11	17
Mean	0.322	50	50	12	15
Types 4 and 1 (primary degree versus primary/lower secondary)					
10	0.802	52	48	4	9
25	0.765	46	54	17	9
50	0.718	55	45	17	15
75	0.726	29	71	13	10
90	0.646	33	67	9	16
Mean	0.737	49	51	10	14
Types 3 and 2 (non degree versus complete secondary)					
10	0.07	86	13	49	31
25	0.094	37	63	21	15
50	0.084	76	24	36	26
75	0.105	42	57	24	12
90	0.099	40	60	26	13
Mean	0.091	33	67	20	14
Types 3 and 1 (non degree versus primary/lower secondary)					
10	0.486	44	56	20	13
25	0.501	29	71	12	8
50	0.494	45	55	17	14
75	0.506	41	59	14	8
90	0.437	44	56	17	9
Mean	0.506	34	66	13	9
Types 2 and 1 (complete secondary versus primary/lower secondary)					
10	0.416	31	69	4	6
25	0.407	36	64	10	9
50	0.41	38	62	11	7
75	0.402	37	63	17	4
90	0.338	49	51	12	6
Mean	0.415	39	61	11	8

Table 7 Quantile decompositions, reading

Quantile	Total test score gap	Explained (percentage)	Unexplained (percentage)	Income (percentage)	Books (percentage)
Types 4 and 3 (primary degree versus non-degree)					
10	0.35	63	37	16	15
25	0.391	43	57	10	11
50	0.401	41	59	7	9
75	0.318	38	62	11	9
90	0.222	47	53	18	11
Mean	0.331	41	59	9	11
Types 4 and 2 (primary degree versus complete secondary)					
10	0.485	68	32	18	20
25	0.511	51	49	12	17
50	0.54	45	54	9	14
75	0.471	40	60	12	12
90	0.377	44	56	18	12
Mean	0.491	40	60	9	14
Types 4 and 1 (primary degree versus primary/lower secondary)					
10	0.798	79	21	19	24
25	0.899	54	46	12	19
50	0.906	49	51	9	16
75	0.785	42	58	13	14
90	0.657	44	57	17	14
Mean	0.856	46	54	10	18
Types 3 and 2 (non degree versus complete secondary)					
10	0.135	61	39	39	27
25	0.12	68	33	26	30
50	0.138	51	49	16	27
75	0.153	40	60	12	27
90	0.155	25	75	5	19
Mean	0.159	41	59	9	24
Types 3 and 1 (non degree versus primary/lower secondary)					
10	0.448	67	33	34	25
25	0.508	63	37	18	21
50	0.504	58	42	13	22
75	0.467	54	46	11	27
90	0.435	36	64	5	21
Mean	0.524	53	47	10	25
Types 2 and 1 (complete secondary versus primary/lower secondary)					
10	0.313	47	53	20	12
25	0.388	43	56	21	11
50	0.366	59	41	22	16
75	0.314	70	30	25	23
90	0.28	63	37	14	18
Mean	0.365	57	43	20	16

4 Results

We now present the results of our analysis. First, we look at the pairwise gaps between the groups. They are presented for all pairwise gaps, and given we have four levels of education (our four "types"), this amounts to six pairwise gaps. We present results for the gaps at the 10$^{\text{th}}$, 25$^{\text{th}}$, 50$^{\text{th}}$, 75$^{\text{th}}$ and 90$^{\text{th}}$ quantiles and also at the mean. The presence of these gaps suggest that inequality of opportunity is present.

We first examine the traditional BO decomposition, evaluated at the mean. Thus for example, in Table 6, looking at the gap evaluated at the mean between groups 4 and 3 (those whose mothers had a primary degree versus those with a post-school but non-degree qualification respectively), we see that for maths the average gap in test scores is 0.231 (about a quarter of a standard deviation). Furthermore, we can see that about half of this gap can be accounted for by differences in observable characteristics between the two groups. The other half arises either from differences in unobserved characteristics or else differences in the returns (in the education production function) to the observed characteristics.

Given that we have six pairwise gaps for both maths and reading, in terms of the detailed decompositions it is possible that we will observe a number of variables which will be statistically significant in the decompositions, with this significance simply reflecting type I errors. Thus, we only show the part of the explained gap accounted for by two specific characteristics, income and the total number of children's books in the study child's house as

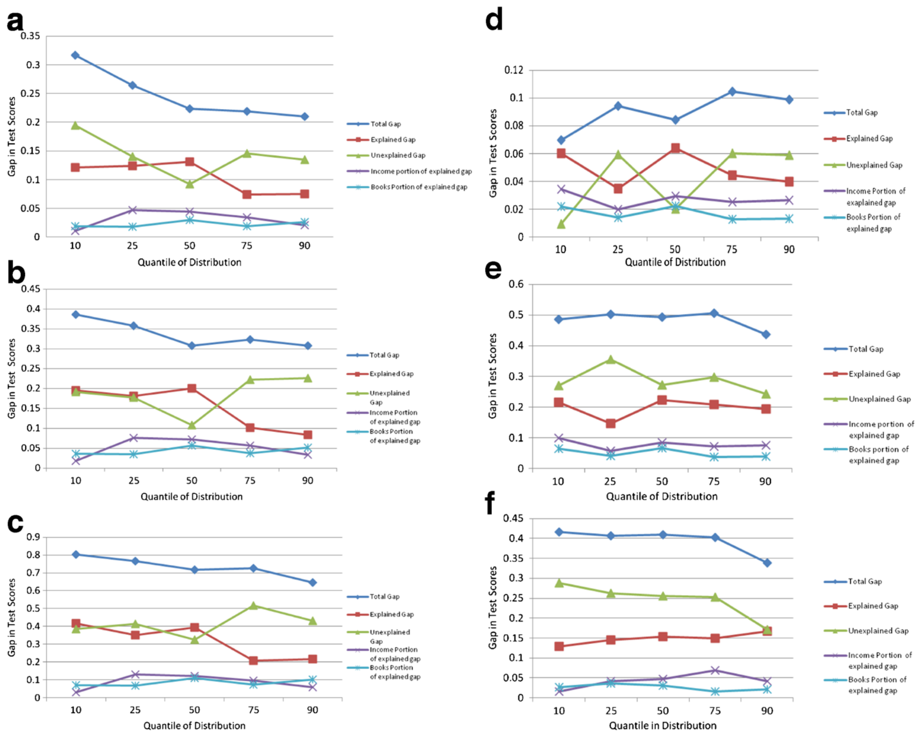

Fig. 2 **a** Quantile decompositions, groups 4 and 3, maths. **b** Quantile Decompositions, groups 4 and 2, maths. **c** Quantile decompositions, groups 4 and 1, maths. **d** Quantile Decompositions, groups 3 and 2, maths. **e** Quantile decompositions, groups 3 and 1, maths. **f** Quantile decompositions, groups 2 and 1, maths

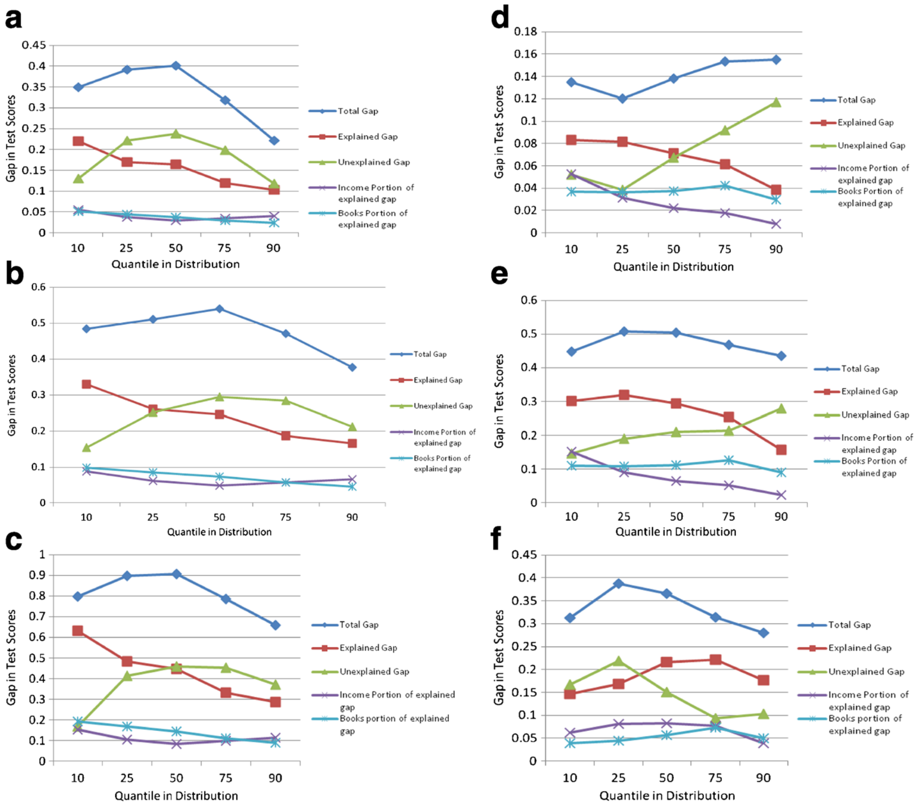

Fig. 3 **a** Quantile decompositions, groups 4 and 3, reading. **b** Quantile decompositions, groups 4 and 2, reading. **c** Quantile decompositions, groups 4 and 1, reading. **d** Quantile decompositions, groups 3 and 2, reading. **e** Quantile decompositions, groups 3 and 1, reading. **f** Quantile decompositions, groups 2 and 1, reading

these variables consistently show up in virtually all detailed decompositions as having statistically significant associations with both maths and reading scores.[8] We thus see that for our example here, these two characteristics account for just under half of the explained difference (the detailed decompositions are available in the Online Appendix, see Tables A13 and A14).

We also show the decomposition of the pairwise gaps evaluated at different quantiles of the distribution. Staying with our example of the gap in maths scores between groups 4 and 3, we see that the gap evaluated at the 10th percentile is almost 0.32, while at the 90th percentile it is down to about 0.21. Thus the gap, and hence inequality of opportunity appears to decline slightly as we move from lower scoring to higher scoring children. The decomposition between explained and unexplained factors is not uniform across the distribution, with a greater role for explained, observable factors in the middle of the distribution, compared to the tails.

[8] For international evidence on the importance of the latter factor in terms of children's educational achievements, see Evans et al. (2010) and Chiu and Chow (2010).

The remainder of Table 6 provides similar results for the other pairwise gaps, while Table 7 provides analogous results for reading. We also present these results visually, in Figs. 2 and 3. In Fig. 2a, which corresponds to the specific set of results we have just discussed, the vertical axis shows the gap in scores, while the horizontal axis corresponds to different quantiles of the distribution. Thus again, we have the total gap, the explained and unexplained portions of it, and the part of the explained portion accounted for by income and books.

Overall, looking at Tables 6 and 7, and Figs. 2 and 3, it seems fair to say that the results show a considerable degree of uniformity across maths and reading and that within each pairwise comparison, the gaps are generally similar across the distribution (with the only exception being the case of reading for type 4 versus the rest). In all pairwise cases, gaps are evident and in the case of the gap between the highest and lowest levels of maternal education, the gap can be quite substantial, indicating the presence of inequality of opportunity. About 30–50% of the pairwise gaps are accounted for by differences in observable circumstances, and within that portion accounted for by differences in observables circumstances, about 50% of the difference arises from differences in income and books within the house.

5 Discussion and conclusion

While by their nature decompositions can be viewed as a sophisticated method of carrying out initial analysis, do the findings here point towards any potential policy conclusions? In commencing this discussion, it is vital to bear in mind that the results we present show statistical associations between our outcome variable and a variety of characteristics and do not address issues of endogeneity and simultaneity, so causality cannot be inferred. In this respect, the discussion which follows should be regarded as preliminary and suggestive. Nevertheless, where consistent statistical associations are found, this may give useful pointers as to where further research into definite policy recommendations might be directed.

As outlined above, given the large number of decompositions which we carry out, it is inevitable that many variables will, on occasion, show up as contributing in a statistically significant way to the explained gap. However, the three factors which show up most consistently are the age of the principal carer, equivalised household income and the number of children's books in the house, with the latter two factors accounting for around 50% of the explained gap. As is ever the case with decompositions of this nature, it is important to be aware of issues regarding endogeneity and the likelihood of simultaneity. Thus it is possible that having a large number of books available in a house improves a child's reading skills. However, it is also possible that the presence of such a large number of books reflects a child's innate interest in and/or aptitude for reading (or indeed that there is a third, unobserved, factor affecting both). However, given that an association has been found in other studies between the number of *adult* books in a house and child educational outcomes it may be the case that at least some of this association reflects a causal effect, in the sense that the number of books influences what Evans et al. (2010) refer to as the degree of "family scholarly culture" present in a house.[9] In the case of principal carer's age and equivalised income, reverse causality with child test scores at age nine seems less plausible and so simultaneity may not be an issue.

[9] It is also interesting to note that books do not appear to have a greater impact when the principal carer has a higher level of education. The inclusion of an interaction term between books and education is insignificant for maths scores and is barely significant (p value of 0.072) and with a small coefficient for reading scores.

It is noticeable that these three characteristics could be regarded as “home” rather than “school” characteristics. Thus it does not appear to be the case that the gaps in test scores arise from differences in school resources, at least not in those school resources which can be directly observed. Thus, policy initiatives which might be explored would include greater access to availability of reading material (or perhaps other educational material or dimensions of scholarly culture). This could be achieved by direct provision of books, perhaps through the schooling system, or via an enhanced public library system.

The other two consistently significant factors are age of principal carer and income. We should note that family income will be greatly influenced by income of other household members (most notably the father, for those households where a father is present). Since fathers’ income is likely to be highly correlated with their education, it is possible that family income is picking up at least partly the influence of fathers education.[10] Possible policies in the area of income which may be worth exploring are the adequacy of current subsidies and grants which assist parents in purchasing educational resources.

A role for income also suggests intergenerational forces that may be at work which exacerbate inequalities. If test scores are influenced by family income, then, assuming that such test scores are good predictors of income for the next generation, this will act as an impediment for children from poorer backgrounds to have high incomes later in life. Exploration of interactions between income and education also suggests that the effect of income on test scores may be greater when the principal carer has lower levels of education, suggesting that a reduction in income inequality in this generation may have positive implications for the next generation.

The age of principal carer is also positively associated with tests scores. This is after controlling for income, education, lone parent status and the presence of younger and older siblings, all factors which might be expected to be correlated with age. Thus, this positive association may simply reflect the fact that older parents have a greater set of parenting skills and experience, consistent with some positive returns (in terms of childhood educational outcomes) to delaying having children (up to a point only of course). Note that while it might be expected that maternal age will be positively correlated with level of education, a general regression of the whole sample shows independent effects of both age and education.

Clearly in any analysis of this nature, the potential role of omitted variables must be considered. Amongst the most important of these are genetic factors. Genetically inherited traits can be regarded as “brute luck” (Dworkin 1981; Ferreira and Peragine 2015). However, unlike social background, where there is general agreement that this is a source of inequality for which some form of compensation should be provided, there is no such universal agreement with respect to inherited traits and it is problematic to see how compensation could be carried out in practice.

If it is also believed that non-cognitive skills such as patience/work ethic etc are inherited as well as cognitive skills, then the distinction between inherited traits and preferences may become blurred (e.g. if someone works very long hours and receives monetary rewards for it, is this a case of a preference or an inherited trait?). When such inherited traits are difficult to measure (as is usually the case), the default option may then be to regard them as preferences rather than circumstances. That would seem to be the case especially when

[10]In preliminary versions of this work we considered the possibility of also using fathers education as a “type”. However, information on fathers education was missing for many (presumably non-random) observations and this would also have implied 28 different pairwise comaprisons.

types are not defined on the basis of that inherited trait, even though it may be highly correlated with the circumstance which is used as the basis for definition of type. It may also be desirable to make a distinction between endowed talent (which presumably is a circumstance) and acquired talent (which could be regarded as arising from effort). But in practice this distinction may be very hard to make, as genes may be influenced by environment, and part of the environment may arise from choices/effort. And in any event, as we discussed above, the concept of effort, as usually applied in the inequality of opportunity literature, does not seem appropriate when dealing with nine-year olds.

Are there any general observations which can be made regarding the unexplained part of the gap? Unlike the case of the explained portion, where variables such as income and the number of books consistently appeared as statistically significant explanatory variables, there appears to be little such pattern in the unexplained component.

In conclusion, this paper has examined inequality of opportunity in education outcomes for nine-year olds in Ireland via quantile decompositions. Four "types" were identified (via the level of maternal education) and pairwise decompositions were carried out at selected quantiles. Consistent with the inequality of opportunity approach, each type reflected a circumstance which was outside the control of the nine-year olds i.e. their mothers education level. Quantile decompositions provided evidence on the factors lying behind the inequality at different points in the distribution, apart from just the mean.

The principal advantage of approaching inequality of opportunity from this perspective is that the detailed decomposition provides some evidence of the characteristics which are associated with (part of) the gap in test scores and which indicate potential policy areas. The results here suggest a consistent role for the number of books in a house, income and maternal age. While it may be difficult to directly affect the latter factor, policy initiatives to address the number of books available to a child, and indeed the resources in a household which support education in general, may be worthy of further exploration.

In our discussion of policy conclusions, we have essentially been assuming, in line with most of the inequality of opportunity literature, that type is exogenous. This is also typically the case in decomposition exercises where the population is usually partitioned along a dimension which is considered exogenous, such as race or gender. While type is clearly beyond the control of the study children, it is a choice variable to some degree for the mothers (although it is likely to be a choice made before they have children and it is arguable that the future implications for children's education achievements may not be a major factor in education decisions which are made during teen and early adult years). Nevertheless, as a general policy it could be reasonably expected that greater equality of education levels amongst mothers would lead to a reduction in inequality of opportunity. However, it also seems reasonable that such a policy should be viewed as more long-term. Given existing differences in type, the decompositions here do point to a menu of other policies which could be considered to address inequality of education.

Acknowledgments I am grateful to Donal O Neill, Barry Reilly, Chris Whelan, Kevin Denny and participants at a Geary Institute seminar for useful comments and discussion. I would also like to thank the editor and three anonymous referees for very helpful comments and suggestions. The usual disclaimer applies.

References

Ashenfelter, O., Harmon, C., Oosterbeek, H.: A review of estimates of the schooling-earning relationship, with tests for publication bias. Labour Econ. **6**, 453–470 (1999)

Baltagi, B., Ghosh, P.K.: Replication of unconditional quantile regressions by firpo, fortin and lemieux (2009). J. Appl. Econ., Article first published online: 14 JUL. https://doi.org/10.1002/jae.2477 (2015)
Blinder, A.: Wage discrimination: reduced form and structural estimates. J. Hum. Resour. **8**, 436–455 (1973)
Björklund, A., Salvanes, K.: Education and family background: Mechanisms and policies. IZA working papers. DP, 5002 (2010)
Breen, R., Goldthorpe, J.: Class inequality and meritocracy: a critique of saunders and an alternative analysis. Br. J. Sociol. **50**, 1–27 (1999)
Breen, R., Goldthorpe, J.: Class, mobility and merit: the experience of two british birth cohorts. Eur. Sociol. Rev. **17**, 81–101 (2001)
Chiu, M., Chow, B.: Culture, motivation and reading achievement: high school students in 41 countries. Learn. Individ. Differ. **20**, 579–592 (2010)
Clotfelter, C., Ladd, H., Vigdor, J.: The academic achievement gap in grades 3 to 8. Rev. Econ. Stat. **91**, 398–419 (2009)
Cunha, F., Heckman, J.: The technology of skill formation. National bureau of economic research, WP 12840 (2007)
Cutler, D., Lleras-Muney, A.: Understanding differences in health behaviours by education. J. Health Econ. **29**, 1–28 (2010)
Denny, K.: What did abolishing university fees in Ireland do? UCD centre for economic research working paper WP10/17 (2010)
Dickson, M., Harmon, C.: Economic returns to education: what we know, what we don't know and where are we going – some brief pointers. Econ. Educ. Rev. **30**, 1118–1122 (2011)
Doris, A., Neill, D., Sweetman, O.: Gender, single sex schooling and maths achievement. Econ. Educ. Rev. **35**, 104–119 (2013)
Dworkin, R.: What is inequality? Part 2: equality of resources. Philos. Public Aff. **10**, 283–345 (1981)
Essama-Nssah, B., Lambert, P.: Influence functions for distributional statistics. Society for the study of economic inequality working paper. ECINEQ WP 2011–236 (2011)
Evans, M., Kelley, J., Sikora, J., Treiman, D.: Family scholarly culture and educational success: books and schooling in 27 nations. Research in Social Stratification and Mobility **28**, 171–197 (2010)
Feinstein, L.: Inequality in the early cognitive development of British children in the 1970 cohort. Economica **70**, 73–97 (2003)
Ferreira, F., Peragine, V.: Equality of opportunity: theory and evidence. World bank policy research paper 7217 (2015)
Firpo, S., Fortin, N., Lemieux, T.: Unconditional quantile regressions. Econometrica **77**, 953–973 (2009)
Fortin, N., Lemieux, T., Firpo, S.: Decomposition Methods in Economics. In: Ashenfelter, O., Card, D. (eds.) Handbook of Labor Economics. 1st edn., vol. 4(4). Elsevier (2011)
Hanushek, E.: The economics of schooling: production and efficiency in public schools. J. Econ. Lit. **23**, 1141–1177 (1986)
Haveman, R., Wolfe, B.: Children's prospects and children's policy. J. Econ. Perspect. **7**, 153–174 (1993)
Kanbur, R., Wagstaff, A.: How useful is inequality of opportunity as a policy construct? CEPR discussion paper DP 10508 (2015)
Lord, F.M.: Applications of Item Response Theory to Practical Testing Problems. Lawrence Erlbaum, Hillsdale, NJ (1980)
Murray, A., McCrory, C., Thornton, M., Williams, J., Quail, A., Swords, L., et al.: Growing up in Ireland: national longitudinal study of children: design, implementation and procedure for the child cohort. Office of the Minister for Children and Youth Affairs, Dublin (2011)
Oaxaca, R.: Male-female wage differentials in urban labour markets. Int. Econ. Rev. **14**, 693–709 (1973)
Oreopolous, P., Salvanes, K.: Priceless: the non-pecuniary benefits to schooling. J. Econ. Perspect. **25**, 159–184 (2011)
Porter, S.: Quantile regression: analyzing changes in distributions instead of means. In: Higher Education: Handbook of Theory and Research, pp. 335–381. Springer International Publishing (2015)
Rosa Dias, P., Jones, A.: Giving equality of opportunity a fair innings. Health Econ. **16**, 109–112 (2007)
Romer, J.: Economic development as opportunity equalisation. World bank policy research working paper 6530 (2013)
Williams, J., Greene, S., Doyle, E., et al.: The lives of 9 year olds: growing up in Ireland, national longitudinal study of children (report 1 of the child cohort) Dublin; The Stationery Office (2009)