Bridging the Gaps: Inequalities in Children's Educational Outcomes in Ireland

David Madden
(University College Dublin)

Revised Version

Corresponding Author: David Madden,
School of Economics,
University College Dublin,
Belfield,
Dublin 4,
Ireland.

Phone: 353-1-7168396

Fax: 353-1-2830068

E-Mail: david.madden@ucd.ie

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 ex post 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: Inequality of opportunity, quantile decomposition.

JEL Codes: I24, D63, C21.

Bridging the Gaps: Inequalities in Children's Educational Outcomes in Ireland

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 and Cutler and Lleras-Muney, 2010, with respect to health). 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 is that arbitrary circumstances of background should not act as an impediment to the acquisition of education. 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 (conditional upon them expending the same level of "effort"). What each individual ultimately makes of the educational opportunities presented to them in terms of effort expended may be regarded as a private concern, but from the point of view of society, it seems desirable 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.

Following on from this, equality of opportunity in education would thus seem to be a worthy policy goal. The analysis of inequality of opportunity (as opposed to inequality of outcome) has enjoyed a revival in recent years and this approach has been applied to areas such as education and health as well as income (e.g. Dworkin, 1981, Gamboa and Waltenberg, 2012, Ferreira and Gignoux, 2011a, Romer, 1998, 2002, Rosa Dias and Jones, 2007, Rosa Dias, 2009). While the precise definitions and approaches to measuring inequality of opportunity

may differ, in all cases a clear distinction is made between what may be regarded as "fair" and "unfair" sources of inequality. In some cases the terms ethically defensible and indefensible have been used. For example, what are sometimes labelled as "circumstances" such as parental socio-economic outcomes are seen as unfair sources of inequality, whereas inequality arising from factors such as effort or lifestyles may be seen as fair. This concept can be applied in a straightforward manner to education. Suppose we observe inequality in educational outcomes, then if someone has a low educational achievement owing to background circumstances over which they had no control, this is deemed ethically offensive. But if the inequality and accompanying low outcome arises because they consciously chose not to apply effort then this is regarded as ethically defensible. Precisely which factors should be regarded as circumstances and which as effort is clearly open to interpretation and there may also be purely random or luck factors which do not comfortably fit into either the circumstances or effort category.[1]

Lying behind this approach, and critical to the measurement of inequality of opportunity in education, is the idea of gaps in educational outcomes between groups of people who differ in arbitrary circumstances, or at least in circumstances which should not affect these outcomes, conditional upon expending the same level of effort. The analysis of gaps in outcomes between different population groups is also the subject matter of a different strand in the economics literature, that arising from the approach of Blinder (1973) and Oaxaca (1973), and extended in recent years by Dinardo, Fortin and Lemieux (1996), Macahado and Mata (2005) and Melly (2005) amongst others.[2] This approach decomposes the gap in an outcome between two well-defined groups into that part arising from characteristics and that part

[1] For example, genetically inherited traits can be regarded as "brute luck" (Dworkin, 1981; Ferreira and Peragine, 2015). We discuss the part played by genetic factors in more detail in our conclusions section.

[2] For a recent survey, see Fortin et al (2011).

arising from the return to characteristics. Although typically applied to gaps in wages, the approach can be applied in other contexts, including education.

The contribution of this paper is to show how the Blinder-Oaxaca (henceforth BO) methodology can be applied to the analysis of inequality of opportunity in a way that provides useful insights and potentially enriches our understanding of such inequality. The application in question examines gaps in educational outcomes for a nationally representative sample of nine year old children, where the children are partitioned into different groups on the basis of the educational level of their mothers. The analysis of inequality of opportunity for this group is of particular interest, since if such inequality is observed at such a young age, it may have persistent lifetime effects. However, the use of a sample of children also raises other issues, particularly with respect to the role of effort. Whether and how to interpret effort (which is central to the measure of inequality of opportunity) in the context of children is open to debate, and we discuss this in some detail below.

The remainder of the paper is laid out as follows: in section 2 we discuss the measurement of inequality of opportunity in more detail, with particular emphasis placed on the approach of John Romer which we adopt here. We also discuss how this can be complemented by the BO methodology and address the important issue of the interpretation of effort in the context of children. In section 3 we discuss our data and present the formal model of decomposition. In section 4 we present results while in section 5 we provide some discussion and concluding comments.

2. Inequality of Opportunity and the BO Decomposition

The recent literature on inequality of opportunity regards inequality in outcomes as having both fair and unfair sources. Differences in outcomes arising from circumstances which are beyond the control of an individual and for which they could not be held responsible may be

regarded as an unfair source of inequality. Correspondingly a difference in outcome arising from deliberate choices regarding factors such as effort and lifestyle are viewed as "fair" sources of inequality. A key issue then becomes identification of circumstances and effort and distinguishing between the two bearing in mind that they may simultaneously influence each other.

In all approaches to measuring inequality of opportunity the population is divided into types whereby each type represents a particular combination of circumstances. Following this, there are a variety of approaches which can be taken to measuring inequality of opportunity and a recent comprehensive survey can be found in Ramos and Van de gaer (2012). What they term the direct way of measuring Inequality of Opportunity is to estimate the degree of inequality in a counterfactual distribution where inequalities due to effort have been removed and what remains is simply inequality arising from circumstances. This has been labelled the *ex ante* approach (see Ramos and Van de gaer, 2012) and is an issue we hope to pursue in future research. [3]

In this paper we employ what is known as the *ex post* approach to identifying inequality of opportunity. It is not entirely accurate to state that we will be measuring *overall* inequality of opportunity, since our approach will be to examine pairwise gaps i.e. inequality of opportunity between two groups or "types". An overall measure of inequality of opportunity would require some form of summation of the pairwise gaps. It is not entirely clear what form of summation that would entail and we leave that for further research (for a possible approach, see Checchi and Peragine, 2010).

[3] For a recent discussion of *ex ante* and *ex post* approaches to measuring inequality of opportunity see Fleurbaey and Peragine (2013).

Our approach relies upon the Romer Identification Assumption in order to identify levels of effort. Following the notation in Romer (2013) suppose that the population is divided into T mutually exclusive and exhaustive types, indicated by $t=1,...,T$, where types are defined by circumstances. In the limit we could define type on the basis of all observable factors.

However, defining types on the basis of all possible combinations of observable factors would lead to an impossibly high number of types and there would inevitably be many types with zero observations. In practice, the precise number of types is left to the choice of the analyst but the types should be meaningful in the sense that each should have a sufficiently different set of circumstances so that they can be realistically regarded as providing a different opportunity set for each type. The number of types also needs to be sufficiently small to enable coherent analysis.

In our application the particular circumstance will be the education level of the mother. This has the advantage that it provides a compact number of types (4) where each type appears to offer a different set of circumstances and where each type might reasonably be regarded as having a different opportunity set. Maternal education level also has the advantage that it is fairly straightforward and accurate to measure, unlike other possible measures such as income or social class.

Thus if we let the outcome in question (as explained in more detail below in our application it will be child test scores) be denoted by y, then the average outcome for type t will depend upon the application of effort, e. It is assumed that the outcome is monotonically increasing in effort, e and thus $y^t=y^t(e)$. Note that RIA rests critically upon this assumption, however it seems to be a reasonable assumption for many applications.

For each type t there will be a distribution of effort, which can be denoted as $G^t(e)$. The Romer Identification Assumption then identifies levels of effort by the rank in the

distribution of effort $G^t(e)$. We then have the function $v^t(\theta)=y^t((G^t)^{-1}(\theta))$ which gives the average value of y for people at the θth quantile of the distribution of effort of their type. Note that this assumption does not assume that there is no relationship between type and effort. It is quite possible (indeed in many cases quite plausible) that effort levels will vary by type, and so the same quantile in the distribution of effort for two types will correspond to a different absolute level of effort. However, this could be seen as a strength of this approach rather than a weakness, since a condition of equality of distribution of effort by type would be extremely restrictive and counter-intuitive.

However, it is not necessary to observe the distribution of effort for each type. If y is montonically increasing in effort for each type, then rank by effort (for each type) will correspond exactly to rank by outcome and we can simply observe $v^t(\theta)$ directly as the level of y for type t at the θth quantile of the distribution of y for that type.

Thus in order to confirm the existence of ex post inequality of opportunity, we can examine the cumulative distribution functions (cdf) for the different types in question. Take for example, figure 1, where we show the cdf for two types (1 and 2) for the outcome variable, y. The cdf on the vertical axis can be used to infer ex post inequality of outcome. If effort is monotonic in outcome, then any particular quantile on the vertical axis can be viewed as representing the same level of effort. We can then read across to find the associated level of outcome for that quantile for each type. Thus we see that for the median level of effort for both type (θ=0.5), the outcome for type 1, $y^1(0.5)$ is less than that for type 2, $y^2(0.5)$, and thus ex post inequality of opportunity is present at that level of effort.

If first order stochastic dominance is observed i.e. the cdf for type 2 lies below that of type 1 for all values of y, then we know that regardless of the level of effort chosen, there will always be inequality of opportunity between the two types. If the cdfs cross, then we have

inequality of opportunity in one direction at one level of effort and inequality of opportunity in the other direction at a different level of effort. As mentioned above, in that case, some form of summary measure across the distributions will be required. Thus comparison of cdfs for different types provide a very useful check for the presence of inequality of opportunity.[4]

To summarise, the RIA permits us to detect the presence of inequality of opportunity by examining outcomes for each type for the same quantile of outcome (thus controlling for effort) and measuring the gap in outcome between types. This leads us naturally to the BO approach. In this approach the overall population is partitioned into types, on the basis of some observed characteristic and a reduced form equation for the outcome under consideration is estimated. Most applications of this approach examine outcomes in labour markets and so the reduced form equation is typically a wage equation, but in our application here it would be an education production function. Since linear regressions hold exactly at the mean, BO showed that for a two-way partition of the population, 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). Subsequent extensions showed how the decomposition can be carried out at other parts of the distribution via quantile regression and this is where the approach complements the measurement of inequality of opportunity. By carrying out the decomposition at a given quantile we are controlling for the level of effort, and the BO decomposition then permits a richer analysis of inequality of opportunity.

[4] It is also worth noting that the area between the cdfs will provide a summary of the difference in mean outcomes between the groups. I am grateful to an anonymous referee for this observation.

The BO decomposition evaluated at the mean is also very informative, but it should be borne in mind that the mean outcome may be at a different quantile for different types, and thus according to the RIA, they could be regarded as expending different levels of effort.

However, with quantile decomposition, the gap at each specific quantile has controlled for effort, and this gap reflects ex post inequality of opportunity. The application of the BO and quantile decomposition approach enables a deeper exploration of the factors lying behind inequality of opportunity. In our particular application we can decompose inequality of opportunity in education into two components: the first of these is perhaps most closely related to "standard" inequality of opportunity measures and reflects inequality which arises owing to differences in observable characteristics (the inputs to the production function), whereby different types have different characteristics. This is the "explained" component of the gap, with a further attractive feature of this decomposition being the possibility of breaking down the explained component into the part arising from individual characteristics, such as maternal age, class-size etc.

The second component is the "unexplained" component of the gap and may reflect differences in unobserved characteristics which affect the gap between types and/or different returns (by type) to individual characteristics.[5] Thus even if the average child in type 1 has exactly the same characteristics as the average child in type 2, there may still be a difference in outcomes, as one of the types obtains a better return from a given set of characteristics than the other. In the early Blinder-Oaxaca literature, this was sometimes referred to as "discrimination". However, it could be viewed in the context of this paper as a manifestation of a particular type of inequality of opportunity and indeed seems to be close to inequality of

[5] More specifically, in the regression equation which "explains" outcomes, the differences in the constant term between types reflects differences in unobservables, while differences in the coefficients of the regressors reflects different "returns" to characteristics.

opportunity as discussed in Breen and Goldthorpe (1999, 2001). They maintain that "...children of 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 way of viewing 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. This interpretation might also be applied to the unexplained portion of the gaps evaluated at given quantiles.

The decomposition of inequality of opportunity into characteristics and returns may also be very useful from a policy perspective. If we observe that a given gap in educational outcomes between types arises primarily from characteristics (such as school facilities or teacher training) then the obvious policy response is to provide such characteristics to the disadvantaged group. However, if the gap primarily arises owing to differences in the returns to such characteristics, then it may indicate deeper, structural issues, as well as perhaps a difference in the endowment of unobserved characteristics, either of which may require a different policy response.

Before concluding this section, it is worth considering in more detail, what exactly is meant by effort in the context of children. As outlined above the precise assignment of factors as circumstance or effort variables and the correlation between effort and circumstances pose significant challenges in empirical measurement of inequality of opportunity. But an additional challenge when dealing with children is whether it is legitimate to regard any part of inequality of outcome as arising from differential application of effort. 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 of this issue (as pointed out by Kanbur and Wagstaff, 2015). Roemer has stated that he believes that it is not appropriate to regard anyone under the age of 16 as consciously exerting effort and this position also appears to be held by de Barros et al (2009) when examining various health and educational outcomes for children in a number of Latin American countries. They state: "For children, access [to education, health] defines opportunity because children (unlike adults) cannot be expected to make the efforts to access these basic goods by themselves". Althought de Barros et al then proceed to calculate a measure inequality of opportunity based upon the Duncan dissimilarity index which implicitly assumes that effort is relevant!

If we assume that in the case of children it is not reasonable to assign any portion of inequality to effort, how are we to interpret the part of inequality not accounted for by circumstance? One possible interpretation would be to regard it not as effort expended by children, but instead effort expended by their parents, on the basis that health/education decisions made by younger children are likely to be heavily influenced by their parents. This could be regarded as a form of principal-agent type relationship with decisions (in terms of effort levels) being made by the agent (parent) on behalf of the principal (child). Of course younger children cannot influence the level of effort expended by parents, and so, from the child's perspective, it should be regarded as a circumstance.[6] Yet it may still be useful to distinguish between this particular circumstance (parental effort with respect to child health/education) and other circumstances such as parental education level, parental income etc, although, as is often the case with respect to inequality of opportunity, it can often be difficult to distinguish between the two. In the case of children who have just reached the age

[6] It is debateable point when exactly children reach the age when they can be regarded as responsible for their own actions. The age of majority (in most countries 18) seems to be a reasonable choice.

of majority it is very likely that past parental levels of effort (a circumstance) will have influenced current levels of effort (of the just matured child). These are complex issues where differing positions are held and a detailed discussion is beyond the scope of this paper (see Jusot, Tubeuf and Trannoy, 2013, for a discussion).

To summarise, when measuring inequality of opportunity in the case of children, the issue of what exactly is meant by effort is open to debate, with principle and practice often differing. Nevertheless, regardless of what interpretation is put on that part of inequality which does not arise from circumstances such as parental education/income etc, it still may be useful to make the distinction between circumstances and "effort". The contribution of this paper, which measures and further decomposes inequality of opportunity at different quantiles of "effort", 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.[7]

3. Data, Summary Statistics and Decomposition

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

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

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 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.[8] 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 is 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 and Batty et al, 2007).

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.

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

Where this data is 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. [9]

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 on or before the age of 16. Category 2 is those who completed secondary schooling, thus leaving formal education at around 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.

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 (the differential achievement by gender for maths is explored in more detail in 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).

[9] 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.

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.

Figures 2a and 2b show cumulative density 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. This indicates 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. The horizontal gap between the cdfs for each education type reflects the gap in scores at that quantile. In section 4 we propose to investigate these gaps in more detail via quantile decomposition.

Finally, table 4 shows the summary statistics of the characteristics used in the BO decomposition, 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).

Decomposition

We now explain the decomposition methodology. We first of all very briefly outline the BO decomposition at the mean, and we then describe how decompositions can be carried out at different quantiles of the distribution.

Suppose we have an outcome, y_t (e.g. the Drumcondra score test for maths or reading) for students in type t (t=1,2), and y_t is a linear function of K variables (characteristics), where t indicates maternal education level. We wish to obtain a decomposition of the difference in outcomes between the two types (note that in our application, since we have four types in total, there will be six possible pairwise decomposition).Thus we have

$$y_t = X_t\beta_t + v_{t,}. \quad E(v_t) = 0, t \in \{1,2\}$$

where X represents a vector of characteristics and β is a vector of returns to characteristics (or slope parameters of the relationship, including the intercept). Since $E(v_t) = 0, t \in \{1,2\}$, then following the well-known result of Blinder (1973) and Oaxaca (1973) the total difference in average educational outcome $\Delta_y^{\mu} = \bar{y}_1 - \bar{y}_2$ can be decomposed as follows:

$$\Delta_y^{\mu} = E(y_1) - E(y_2) = E(X_1)\beta_1 - E(X_1)\beta_2 + E(X_1)\beta_2 - E(X_2)\beta_2.$$

$$= (E(X_1)[\beta_1 - \beta_2]) + ([E(X_1) - E(X_2)]\beta_2).$$

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. 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".[10][11]

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.

In the case of a two-way decomposition, this vector can either be the vector of returns for one of the groups, or it could be the vector of returns for the pooled sample, consisting of both groups. The choice here really depends upon the underlying research question which the decomposition is trying to address. If it is believed that the gap in outcomes between two groups arises from undue positive discrimination in favour of one group, then the reference

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

[11] 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.

vector should be that of the lower achieving group. However, if the gap arises owing to undue negative discrimination against one of the groups, then the reference vector should be that of the higher achieving group. This is the approach which is taken in this paper, as it seems more likely that lower achieving groups are being discriminated against. It also seems to be more consistent with the idea that if we were to move towards greater equality in outcomes that it would be a levelling up, rather than a levelling down.[12]

However, we may also be interested in gaps and decompositions at parts of the distribution other than the mean, and of course, given that we are looking at inequality of opportunity using the Romer Identification Assumption then we will wish to examine gaps at different quantiles and not at 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

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

where X_i represents the set of covariates for individual i and β_θ is the coefficient vector for the θth 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

[12] 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

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.

Can we reconstruct the counterfactual unconditional distribution of outcomes $Q_\theta^C = X'_{1i}\beta_{2\theta}$ using estimates from the conditional quantile regressions, and hence carry out a BO type decomposition of the gap evaluated at each quantile? This is straightforward when dealing with the mean, since the law of iterated expectations tells us that $E(y) = E_X(E(y|X))$. Thus the OLS estimate for covariate X_i provides the effect of the covariate on either the conditional or unconditional mean of y. However, the key issue is that 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 θth 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.

Is it possible to move between conditional and unconditional quantiles, so that we can use estimtates from conditional quantile regressions to investigate decompositions of the unconditional distribution at various quantiles? A number of approaches have been suggested, and we choose to follow that of Firpo, Fortin and Lemieux (henceforth FFL, 2009). This has the advantage of avoiding the computationally more intensive approaches of Machado and Mata (2005) and Melly (2005) and 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 a 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-μ. 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-μ+μ=y.

More generally, if *F(y)* is the cumulative distribution of the outcome variable and if *T(.)* is the distributional statistic in question e.g. a quantile, or perhaps the Gini coefficient, 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 θth quantile of the unconditional distribution of the outcome variable and $f_yQ_\theta(y)$ is the density of the marginal distribution of y evaluated at Q_θ (see Essama-Nssah and Lambert, 2011). The RIF is then

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

It is worth noting that in the expression above, apart from the constant terms, $Q_\theta(y)$ and $f_y Q_\theta(y)$, the RIF is an indicator function for whether the outcome variable is smaller than or equal to the quantile value. This can be estimated via a linear probability model for say type 1, and counterfactual proportions for type 2 could then be constructed using type 1's coefficients. The counterfactual proportions can then be inverted back to the counterfactual outcome quantiles and standard decomposition analysis can be applied, including a detailed decomposition (though the omitted category issue remains). As we are specifically interested in the detailed decomposition for the explained part of the decomposition, we choose to use the FFL approach for our quantile decompositions.[13]

4. Results

We now present the results of our analysis. First of all 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 10th, 25th, 50th , 75th and 90th 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, bearing in mind that if we accept that effort can be applied in the context of children that the mean will in all likelihood represent a different level of effort for each type. Thus for example, in table 5, 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) 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 "explained" by differences in

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

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 these variables consistently show up in virtually all detailed decompositions as having statistically significant associations with both maths and reading scores. [14] We thus see that these two characteristics account for just under half of the explained difference.

In our results we also show the decomposition of the pairwise gaps evaluated at different quantiles of the distribution, and these can be interpreted as controlling for effort. 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.

The remainder of table 5 provides similar results for the other pairwise gaps, while table 6 provides analogous results for reading. We also present these results visually, in figures 3 and 4. Thus in figure 3a, which corresponds to the specific set of results we have just

[14] 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).

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.

In terms of the overall results for maths, the pairwise gaps reflect the results in table 3, with substantial gaps for all pairwise comparisons (with the exceptions of those between types 3 and 2, which are more modest). While the gaps do tend to fall as we move to higher quantiles, in truth there is not a lot of variation across the distribution. In general, the unexplained part of the gap slightly exceeds the explained part, except around the median. The fraction of the total gap accounted for by the explained part varies between about one third and one half. In turn, the portion of the explained gap accounted for by income and books is around one half, though it tends to be lower for the 10th quantile.

Turning now to the pairwise gaps for reading, once again the gaps tend to fall slightly with higher quantiles, with the exception of the gap between non-degree qualification and completed secondary school (groups 3 and 2). Contrary to the case with maths, the explained portion of the gap is much closer to the unexplained portion and in the case of pairwise gaps *not* involving type 4, it tends to account for the larger fraction. The portion of the explained gap accounted for by income and books again is around a half in most cases, though somewhat larger in the case of pairwise gaps *not* involving type 4.

Overall, however, even allowing for some differences, it does not seem unfair to say that the results show a considerable degree of uniformity across maths and reading. 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 is quite substantial, thus 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

Our results show evidence of gaps in test scores between children of different "type", where type is defined by the education level of the mother. Since this is clearly beyond the control of the child, these gaps can be seen as evidence of inequality of opportunity. We find that up to half of the gaps are accounted for by observable characteristics, with the other half accounted for by differential returns to these characteristics (by type) or by unobserved characteristics. For that portion of the gap arising from observable characteristics, about half in turn is accounted for by equivalised household income and childrens books in the house.

While by their nature decompositions are in many ways a sophisticated method of carrying out initial analysis, do the findings here point towards any potential policy conclusions? Clearly, before commencing this discussion it is vital to bear in mind that the results we present essentially show statistical associations between our outcome variable and a variety of characteristics. We do not address issues of endogeneity and simultaneity and 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, then 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 mentioned above, ever 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. 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.[15] 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 the simultaneity may not be a problem.

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.

[15] 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.

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.[16] 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 acerbate 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 in having high incomes later in life. Exploration of interaction 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

[16] 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.

be expected that maternal age will be positively correlated with level of education, a general regression of the whole sample shows independent effects for 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, 2014). 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 not such universal agreement with respect to inherited traits. Nozick's (1974) views regarding "self-ownership" implies that people should benefit from inborn traits and it is also 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 very 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 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.

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. By carrying out the decompositions at quantiles it was possible to invoke the Romer identification assumption and essentially control for the level of effort. 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.

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 level amongst mothers would lead to a reduction in inequality of opportunity.

This argument could be made for other applications in the inequality of opportunity literature where type is defined by parental education (e.g. Checci and Peragine, 2010). However, it 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.

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

References

Ashenfelter, O., C. Harmon and H, Oosterbeek: A Review of Estimates of the Schooling-Earning Relationship, with Tests for Publication Bias. Labour Economics. 6, 453-470 (1999).

Baltagi, B. and P.K. Ghosh: Replication of Unconditional Quantile Regressions by Firpo, Fortin and Lemieux (2009). Journal of Applied Econometrics, Article first published online: 14 JUL 2015, DOI: 10.1002/jae.2477.

Batty, G.D., I. Deary, I. Schoon and C. Gale: Mental Ability Across Childhood in Relation to Risk Factors for Premature Mortality in Adult Life: the 1970s British Cohort Study. Journal of Epidemiology and Community Health. 6, 997-1003 (2007).

Blinder, A.: Wage Discrimination: Reduced Form and Structural Estimates. Journal of Human Resources. 8, 436-455 (1973).

Breen, R., and J. Goldthorpe: Class Inequality and Meritocracy: a Critique of Saunders and an Alternative Analysis. British Journal of Sociology. 50, 1-27 (1999).

----------., and ----------: Class, Mobility and Merit: the Experience of Two British Birth Cohorts", European Sociological Review. 17, 81-101 (2001).

Checci, D., and V. Peragine: Inequality of Opportunity in Italy. Journal of Economic Inequality. 8, 429-450 (2010).

Chiu, M., and B. Chow: Culture, Motivation and Reading Achievement: High School Students in 41 Countries. Learning and Individual Differences. 20, 579-592 (2010).

Clotfelter, C., H. Ladd and J. Vigdor: The Academic Achievement Gap in Grades 3 to 8. Review of Economics and Statistics. 91, 398-419 (2009).

Cutler, D. and A. Lleras-Muney: Understanding Differences in Health Behaviours by Education. Journal of Health Economics. 29, 1-28 (2010).

De Barros, R., F. Ferreira, J. Molinas Vega and J. Saavedra Chanduvi: Measuring Inequalities of Opportunity in Latin America and the Caribbean. The World bank. Washington DC (2009).

Dickson, M., and C. Harmon: Economic Returns to Education: What We Know, What We Don't Know and Where Are We Going – Some Brief Pointers. Economics of Education Review. 30, 1118-1122 (2011).

DiNardo, J., N. Fortin and T. Lemieux: Labour Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach. Econometrica. 64, 1001-1044 (1996).

Doris, A., D. O Neill and O. Sweetman: Gender, Single Sex Schooling and Maths Achievement. Economics of Education Review. 35, 104-119 (2013).

Dworkin, R.: What is Inequality? Part 2: Equality of Resources. Philosophy and Public Affairs. 10, 283-345 (1981).

Essama-Nssah, B., and P. Lambert: Influence Functions for Distributional Statistics. Society for the Study of Economic Inequality Working Paper. ECINEQ WP 2011-236 (2011).

Evans, M., J. Kelley, J. Sikora and D. Treiman: 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., and J. Gignoux: The Measurement of Educational Inequality: Achievement and Opportunity. IZA Working Papers. DP 6161 (2011a).

----------, and V. Peragine: Equality of Opportunity: Theory and Evidence. World bank Policy Research Paper 7217 (2015).

Firpo, S., N. Fortin and T. Lemieux: Unconditional Quantile Regressions. Econometrica. 77, 953-973 (2009).

Fleurbaey, M., and V. Peragine: "Ex Ante versus Ex Post Inequality of Opportunity". Economica. 80, 118-130 (2013).

Fortin, N., T. Lemieux and S. Firpo in O. Ashenfelter & D. Card (ed.), 2011. Handbook of Labor Economics, Elsevier, edition 1, volume 4, number 4 (2011).

Gamboa, L., and F. Waltenberg: Inequality of Opportunity for Educational Achievement in Latin America: Evdience from PISA, 2006-2009. Economics of Education Review. 31, 694-708 (2012).

Hanushek, E.: The Economics of Schooling: Production and Efficiency in Public Schools. Journal of Economic Literature. 23, 1141-1177 (1986).

Haveman, R., and B. Wolfe: Children's Prospects and Children's Policy. Journal of Economic Perspectives. 7, 153-174 (1993).

Jusot, F., S. Tubeuf, and A. Trannoy: Circumstances and Efforts: How important is their correlation for the measurement of inequality of opportunity in health? Health Economics 22, 1470-1495 (2013).

Lord, F. M.: Applications of Item Response Theory to Practical Testing Problems. Hillsdale, NJ: Lawrence Erlbaum (1980).

Machado, J., and J. Mata: Counterfactual Decomposition of Changes in Wage Distributions Using Quantile Regression. Journal of Applied Econometrics. 20, 445-465 (2005).

Melly, B.,: Decomposition of Differences in Distributions Using Quantile Regression. Labour Economics. 12, 577-590 (2005).

Murray, A., C. McCrory, M. Thornton, J. Williams, A, Quail, L. Swords et al: Growing Up in Ireland: National Longitudinal Study of Children: Design, Implementation and Procedure for the Child Cohort. Dublin: Office of the Minister for Children and Youth Affairs (2011).

Nozick, R.,: Anarchy, State and Utopia. New York, Basic Books (1974).

Oaxaca, R.: Male-Female Wage Differentials in Urban Labour Markets. International Economic Review. 14, 693-709 (1973).

Ramos, X., and D. Van de gaer: Empirical Approaches to Inequality of Opportunity: Principles, Measures and Evidence. IZA Discussion paper 6672 (2012).

Rosa Dias, P.: Inequality of Opportunity in Health: Evidence from a UK Cohort Study. Health Economics. 18, 1057-1074 (2009).

Rosa Dias, P. and A, Jones: Giving Equality of Opportunity a Fair Innings. Health Economics. 16, 109-112 (2007).

Romer, J: Equality of Opportunity. Harvard University Press, Harvard (1998).

Romer, J: Equality of Opportunity: A Progress Report. Social Choice and Welfare. 19, 455-471 (2002).

Romer, J: Economic Development as Opportunity Equalisation. World Bank Policy Research Working Paper 6530 (2013).

Wagstaff, A. and R. Kanbur: How Useful is Inequality of Opportunity as a Policy Construct? CEPR Discussion Paper DP 10508 (2015).

Williams, J., S. Greene, E. Doyle 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).

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**

Table 2: Summary Drumcondra Logit Scores by Gender (standard deviation in brackets)

	Total	Female	Male
Maths	-0.759 (0.933)	-0.822 (0.879)	-0.699 (0.979)
Reading	0.012 (0.994)	0.015 (0.965)	0.009 (1.020)

Table 3: Mean Scores by Type (SD in brackets)

	Maths	Reading
1. Primary/ Low Sec	-1.121 (0.913)	-0.355 (0.967)
2. Complete Secondary	-0.707 (0.897)	0.009 (0.948)
3. Non-Degree	-0.616 (0.898)	0.169 (0.970)
4. Primary Degree	-0.385 (0.870)	0.500 (0.907)

Table 4 Summary Statistics – Mean and Standard Deviation

Variable	**Full Sample (mean)**	**Full Sample (s.d.)**	**1. Prim Low Sec (mean)**	**1. Prim Low Sec (s.d)**	**2. Sec. Educ (mean)**	**2. Sec Educ (s.d.)**	**3. Non Degree (mean)**	**3. Non Degree (s.d.)**	**4. Degree (mean)**	**4. Degree (s.d.)**
Age	39.01624	*5.533395*	38.10967	*5.827417*	39.14683	*5.416292*	38.69879	*5.424348*	40.59088	*4.986012*
Birthweight	3.512211	*0.621518*	3.43699	*0.631708*	3.536918	*0.61624*	3.559353	*0.580597*	3.542842	*0.641954*
Early birth	0.016702	*0.128161*	0.023986	*0.153065*	0.014413	*0.119212*	0.014671	*0.120263*	0.011107	*0.104829*
Smoker	0.250138	*0.433121*	0.413914	*0.492722*	0.207745	*0.405778*	0.189026	*0.391633*	0.119134	*0.324029*
Smoked while pregnant	0.163104	*0.369485*	0.309954	*0.462652*	0.129213	*0.335506*	0.09003	*0.286301*	0.053963	*0.226003*
Drank while pregnant	0.013986	*0.11744*	0.013937	*0.117274*	0.01422	*0.118422*	0.013563	*0.115697*	0.01396	*0.117355*
Breastfed	0.445675	*0.497073*	0.240222	*0.427382*	0.420079	*0.493675*	0.571808	*0.494948*	0.735327	*0.441272*
Illness	0.104613	*0.306074*	0.133903	*0.340678*	0.093315	*0.290933*	0.100623	*0.300908*	0.082729	*0.275542*
Equivalised Income (log)	9.702217	*0.518019*	9.436605	*0.504554*	9.713515	*0.446158*	9.818277	*0.43969*	10.02422	*0.526439*
Mother Healthy	0.933383	*0.249374*	0.882161	*0.322541*	0.951218	*0.215456*	0.95504	*0.207271*	0.961936	*0.1914*
Trauma	1.493229	*1.300045*	1.505869	*1.339595*	1.412205	*1.267734*	1.591387	*1.31664*	1.555532	*1.275681*
Books in house	4.170796	*1.080595*	3.775404	*1.217242*	4.20866	*1.036112*	4.379503	*0.926399*	4.570027	*0.810674*
Local 1	-12.6579	*2.764782*	-11.9568	*3.063673*	-12.8975	*2.684501*	-12.807	*2.625989*	-13.1982	*2.2471*

Local 2	6.413011	*1.842896*	6.545216	*1.765734*	6.399393	*1.906315*	6.389999	*1.803272*	6.237255	*1.854361*
Mother Working	0.539861	*0.498442*	0.397341	*0.489535*	0.545712	*0.49801*	0.596408	*0.490748*	0.718528	*0.449832*
Class size	26.03057	*6.392365*	24.84783	*7.148572*	26.35415	*6.101057*	26.5402	*6.124644*	26.87375	*5.558825*
Parent/teacher meeting	0.884662	*0.319451*	0.852402	*0.354837*	0.903678	*0.295094*	0.886477	*0.317316*	0.896843	*0.304242*
Engagement	2.861207	*0.870445*	2.853714	*0.879744*	2.856596	*0.884823*	2.869244	*0.837294*	2.876552	*0.854189*
Teacher Experience	12.75341	*11.29348*	12.47041	*11.47349*	13.23762	*11.39785*	12.42015	*11.05629*	12.49754	*10.94713*
Teacher Qualifications	2.416394	*2.96398*	2.211228	*2.74193*	2.398538	*2.995638*	2.401573	*2.971253*	2.822326	*3.208362*
School size	402.3335	*1951.098*	429.4191	*2014.927*	329.3865	*1769.535*	541.9783	*2252.441*	382.7877	*1903.889*
Younger sibling	0.782546	*0.864487*	0.72313	*0.86776*	0.739557	*0.832545*	0.842687	*0.858911*	0.921672	*0.913326*
Older sibling	0.958499	*0.989698*	1.123747	*1.059697*	0.942333	*0.954521*	0.815002	*0.929703*	0.845763	*0.956153*
Partner in household	0.834688	*0.371486*	0.782631	*0.412614*	0.853371	*0.353809*	0.835852	*0.370508*	0.882262	*0.322381*

Table 5: Quantile Decompositions, Maths

Quantile	Total Test Score Gap	Explained	Unexplained	Income	Books
Types 4 and 3 (Primary Degree versus Non-Degree)					
10	0.316	0.122	0.195	0.011	0.019
25	0.264	0.124	0.140	0.047	0.018
50	0.224	0.132	0.092	0.044	0.029
75	0.219	0.074	0.146	0.035	0.019
90	0.210	0.075	0.134	0.021	0.026
Mean	0.231	0.113	0.118	0.026	0.025
Types 4 and 2 (Primary Degree versus Complete Secondary)					
10	0.386	0.195	0.191	0.018	0.036
25	0.358	0.181	0.177	0.076	0.034
50	0.308	0.2	0.108	0.071	0.057
75	0.324	0.102	0.222	0.056	0.037
90	0.308	0.083	0.225	0.034	0.051
Mean	0.322	0.161	0.16	0.039	0.047
Types 4 and 1 (Primary Degree versus Primary/Lower Secondary)					
10	0.802	0.418	0.384	0.031	0.071
25	0.765	0.352	0.414	0.132	0.067
50	0.718	0.393	0.324	0.123	0.111
75	0.726	0.209	0.517	0.097	0.073
90	0.646	0.215	0.431	0.058	0.101
Mean	0.737	0.361	0.376	0.073	0.103

Table 5: Quantile Decompositions, Maths (contd)

Quantile	Total Test Score Gap	Explained	Unexplained	Income	Books
Types 3 and 2 (Non Degree versus Complete Secondary)					
10	0.07	0.06	0.009	0.034	0.022
25	0.094	0.035	0.059	0.02	0.014
50	0.084	0.064	0.02	0.03	0.022
75	0.105	0.044	0.06	0.025	0.013
90	0.099	0.04	0.059	0.026	0.013
Mean	0.091	0.03	0.061	0.018	0.013
Types 3 and 1 (Non Degree versus Primary/Lower Secondary)					
10	0.486	0.215	0.271	0.099	0.065
25	0.501	0.147	0.355	0.058	0.041
50	0.494	0.222	0.272	0.085	0.067
75	0.506	0.209	0.298	0.072	0.038
90	0.437	0.194	0.243	0.076	0.04
Mean	0.506	0.174	0.332	0.067	0.048
Types 2 and 1 (Complete Secondary versus Primary/Lower Secondary)					
10	0.416	0.128	0.288	0.015	0.026
25	0.407	0.145	0.262	0.041	0.037
50	0.41	0.154	0.256	0.047	0.03
75	0.402	0.149	0.252	0.069	0.016
90	0.338	0.167	0.171	0.041	0.021
Mean	0.415	0.162	0.253	0.046	0.032

Table 6: Quantile Decompositions, Reading

Quantile	Total Test Score Gap	Explained	Unexplained	Income	Books
Types 4 and 3 (Primary Degree versus Non-Degree)					
10	0.35	0.22	0.13	0.055	0.051
25	0.391	0.17	0.221	0.038	0.044
50	0.401	0.164	0.237	0.03	0.038
75	0.318	0.12	0.198	0.035	0.029
90	0.222	0.104	0.118	0.041	0.024
Mean	0.331	0.137	0.195	0.029	0.036
Types 4 and 2 (Primary Degree versus Complete Secondary)					
10	0.485	0.33	0.155	0.089	0.098
25	0.511	0.26	0.251	0.061	0.086
50	0.54	0.245	0.294	0.049	0.074
75	0.471	0.187	0.284	0.057	0.057
90	0.377	0.166	0.211	0.066	0.046
Mean	0.491	0.197	0.293	0.043	0.069
Types 4 and 1 (Primary Degree versus Primary/Lower Secondary)					
10	0.798	0.631	0.167	0.153	0.193
25	0.899	0.484	0.414	0.105	0.168
50	0.906	0.446	0.459	0.084	0.145
75	0.785	0.332	0.453	0.099	0.111
90	0.657	0.286	0.372	0.114	0.09
Mean	0.856	0.392	0.464	0.082	0.151

Table 6: Quantile Decompositions, Reading (contd)

Quantile	Total Test Score Gap	Explained	Unexplained	Income	Books
Types 3 and 2 (Non Degree versus Complete Secondary)					
10	0.135	0.083	0.052	0.053	0.037
25	0.12	0.081	0.039	0.031	0.036
50	0.138	0.071	0.067	0.022	0.037
75	0.153	0.061	0.092	0.018	0.042
90	0.155	0.038	0.117	0.008	0.03
Mean	0.159	0.065	0.094	0.014	0.038
Types 3 and 1 (Non Degree versus Primary/Lower Secondary)					
10	0.448	0.301	0.147	0.152	0.11
25	0.508	0.319	0.189	0.091	0.109
50	0.504	0.294	0.21	0.064	0.111
75	0.467	0.254	0.213	0.051	0.127
90	0.435	0.156	0.279	0.022	0.09
Mean	0.524	0.277	0.247	0.052	0.133
Types 2 and 1 (Complete Secondary versus Primary/Lower Secondary)					
10	0.313	0.147	0.167	0.062	0.038
25	0.388	0.168	0.219	0.081	0.044
50	0.366	0.216	0.15	0.082	0.057
75	0.314	0.221	0.093	0.077	0.073
90	0.28	0.177	0.104	0.039	0.05
Mean	0.365	0.209	0.156	0.072	0.059

Figure 1

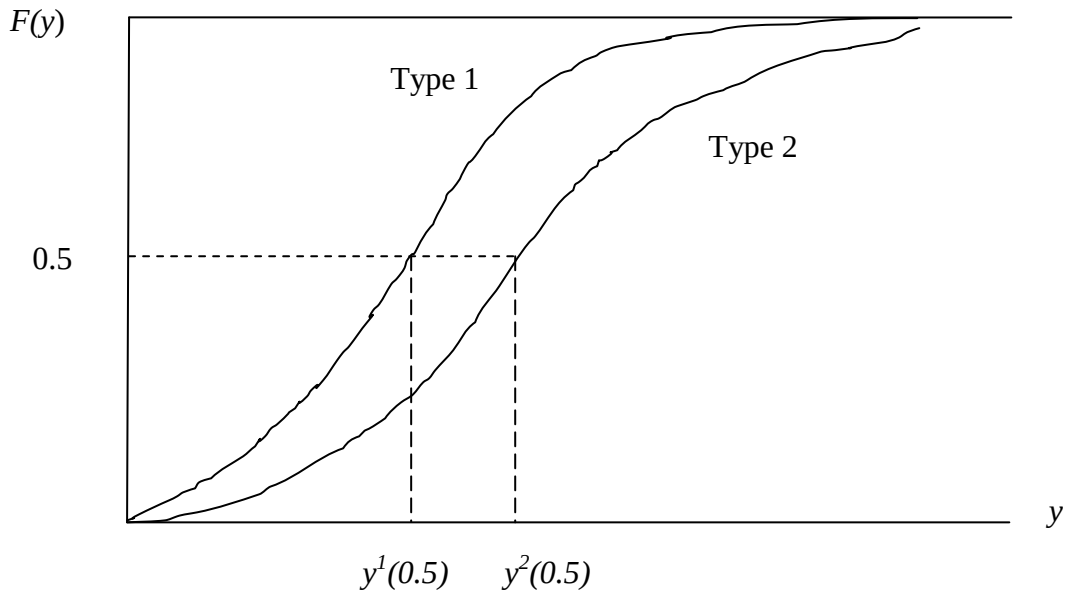

Figure 2a: CDFs Maths Scores by Education

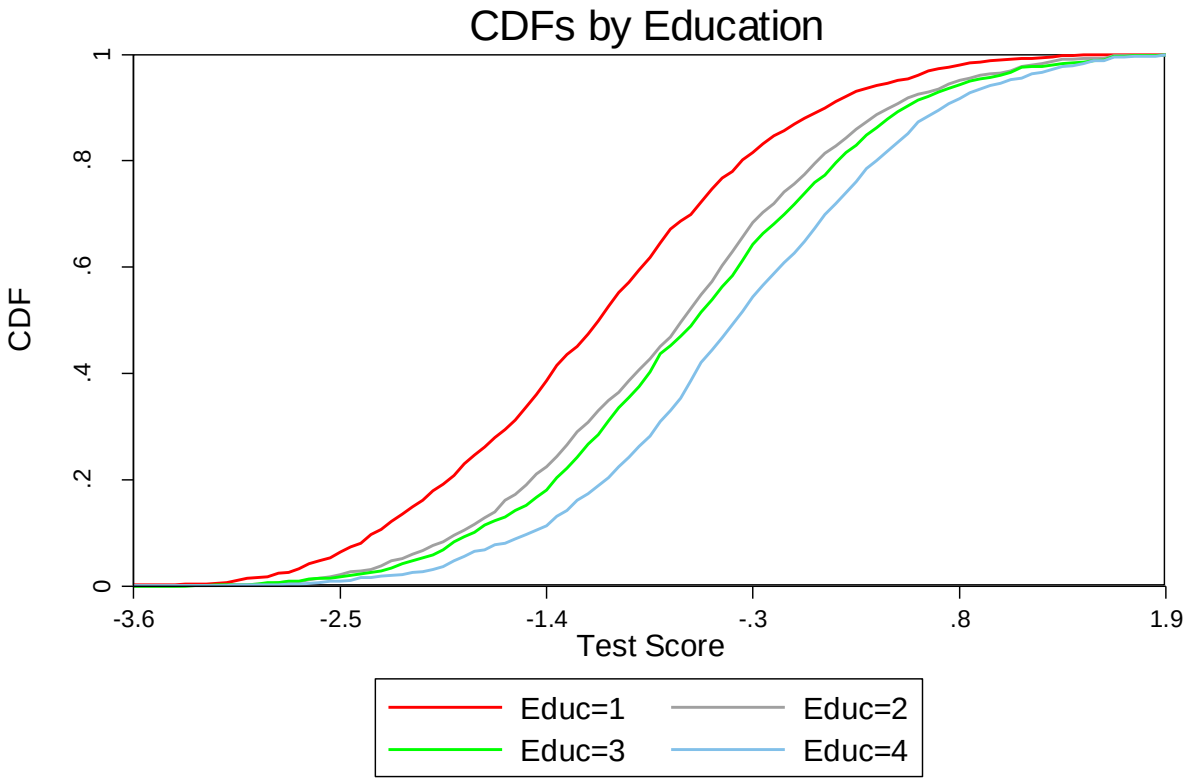

Figure 2b: CDFs Reading Scores by Education

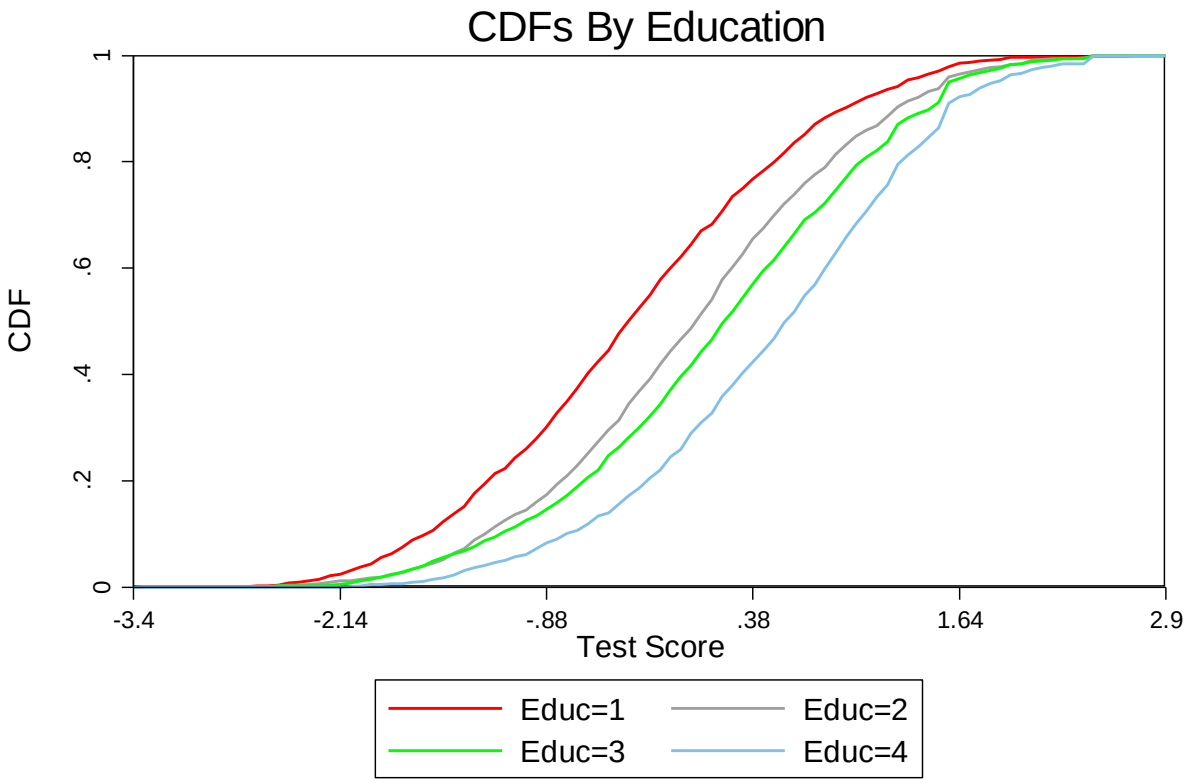

Figure 3a: Quantile Decompositions, Groups 4 and 3, Maths

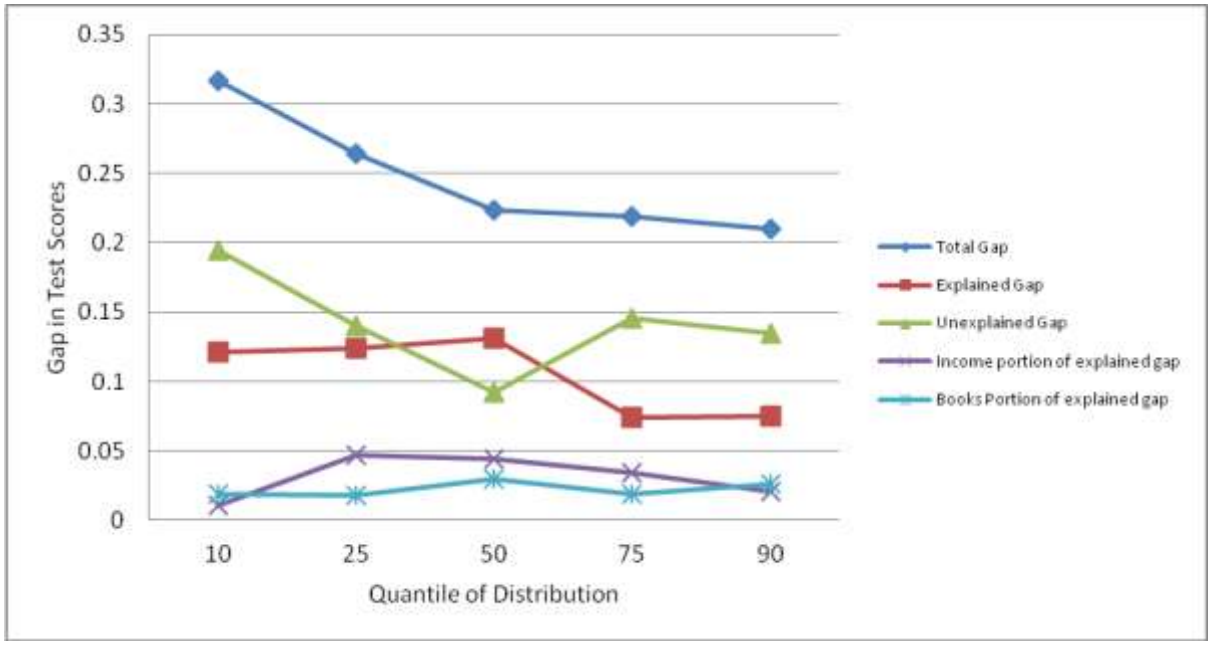

Figure 3b, Quantile Decompositions, Groups 4 and 2, Maths

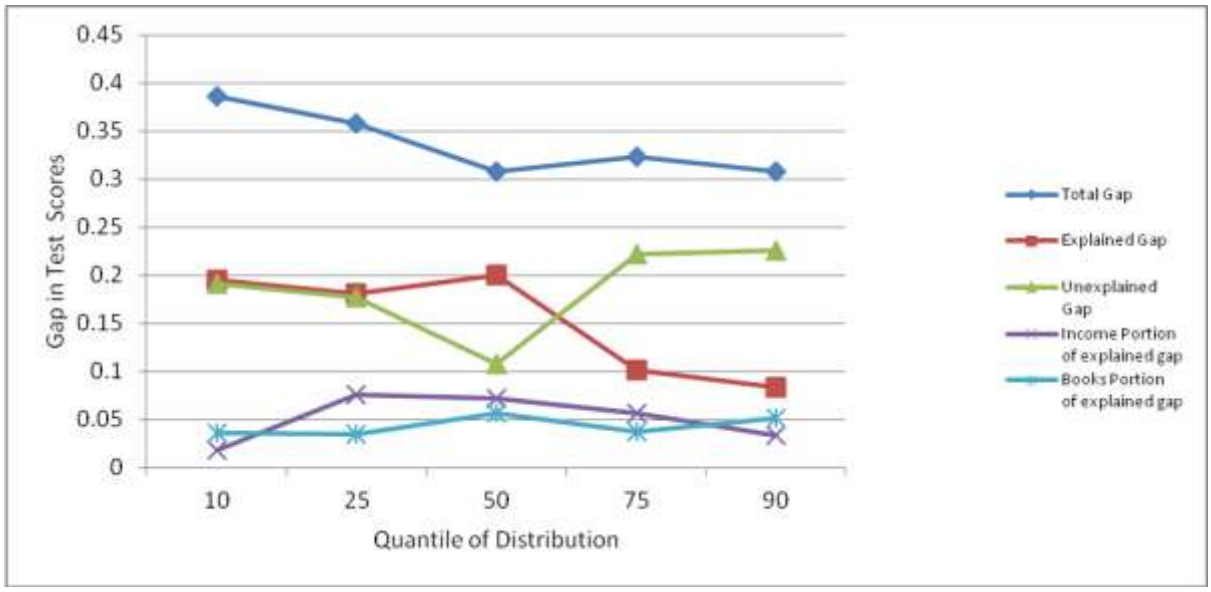

Figure 3c: Quantile Decompositions, Groups 4 and 1, Maths

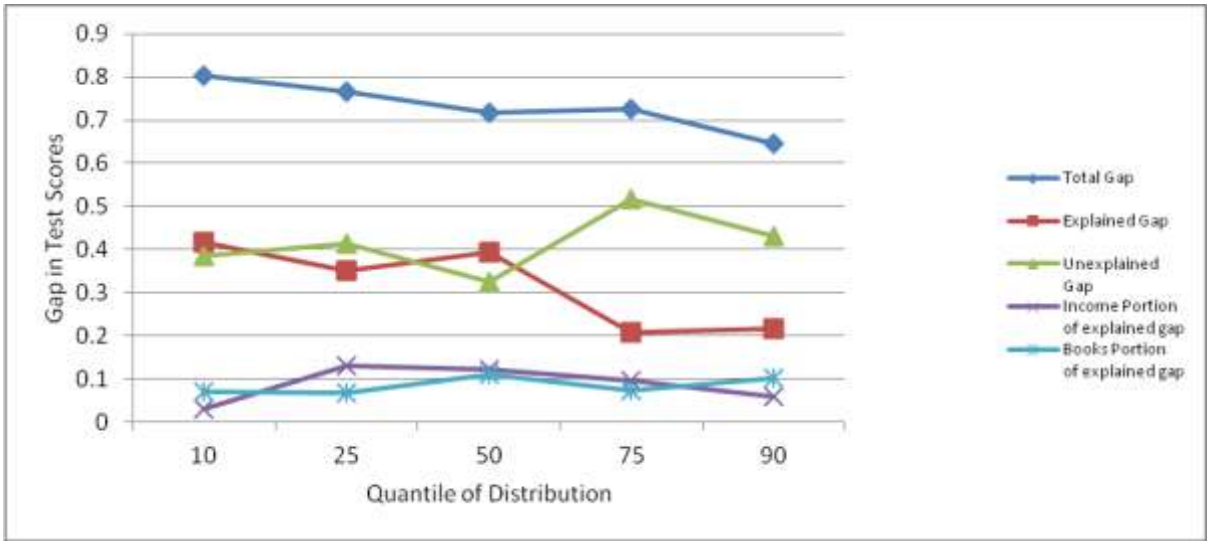

Figure 3d, Quantile Decompositions, Groups 3 and 2, Maths

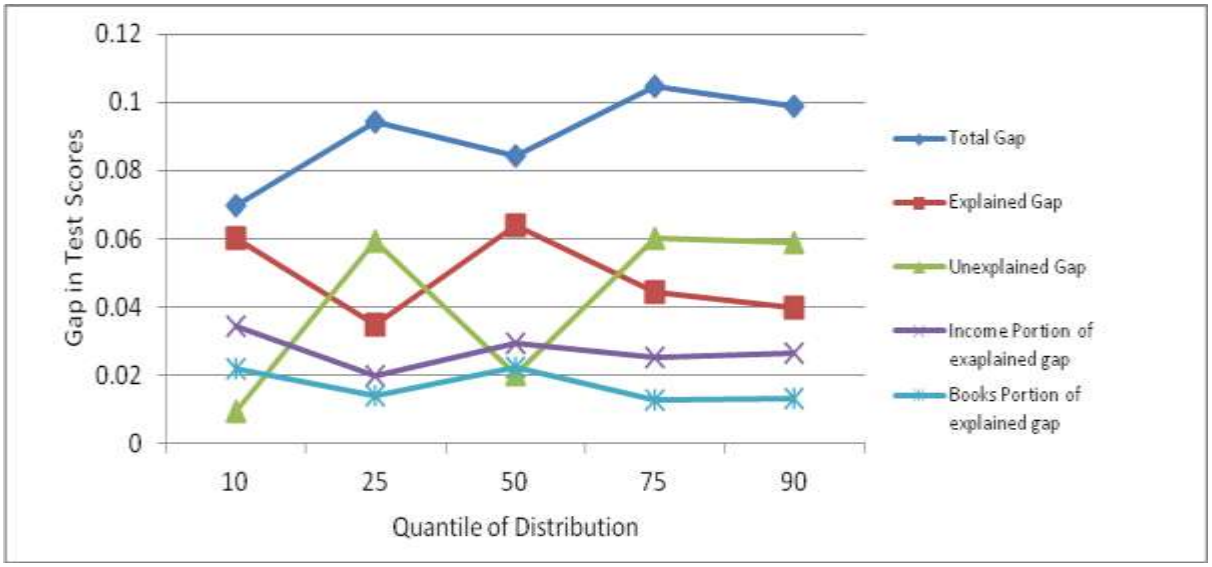

Figure 3e: Quantile Decompositions, Groups 3 and 1, Maths

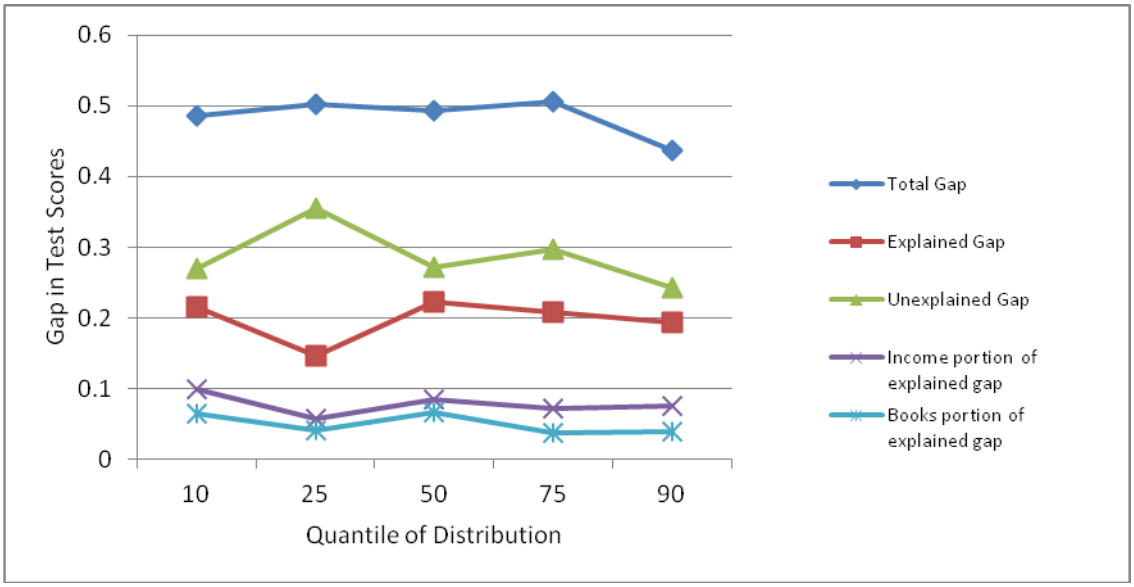

Figure 3f: Quantile Decompositions, Groups 2 and 1, Maths

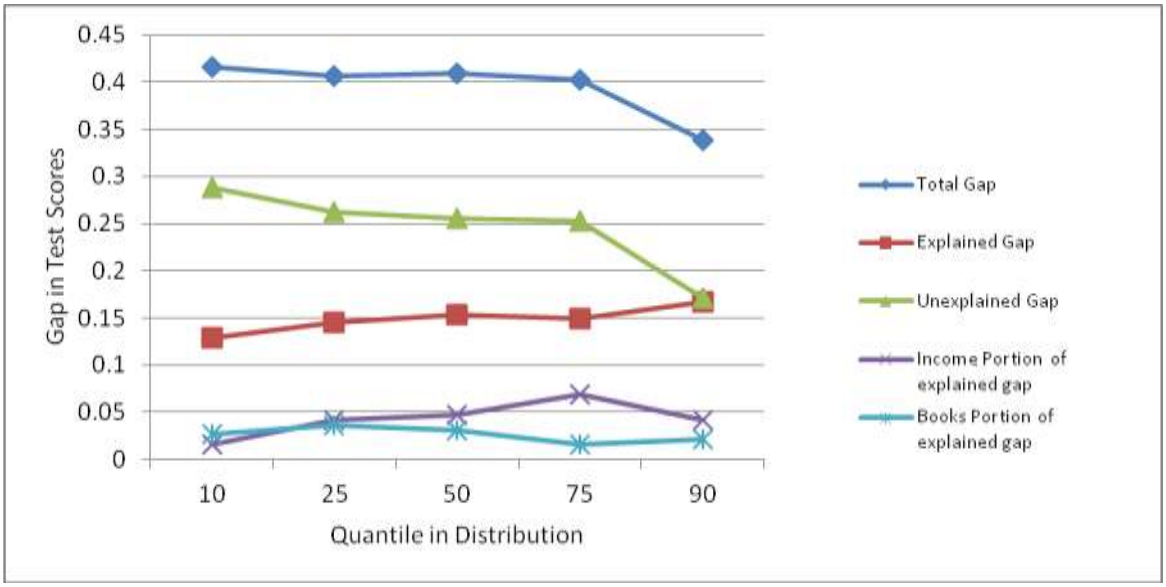

Figure 4a: Quantile Decompositions, Groups 4 and 3, Reading

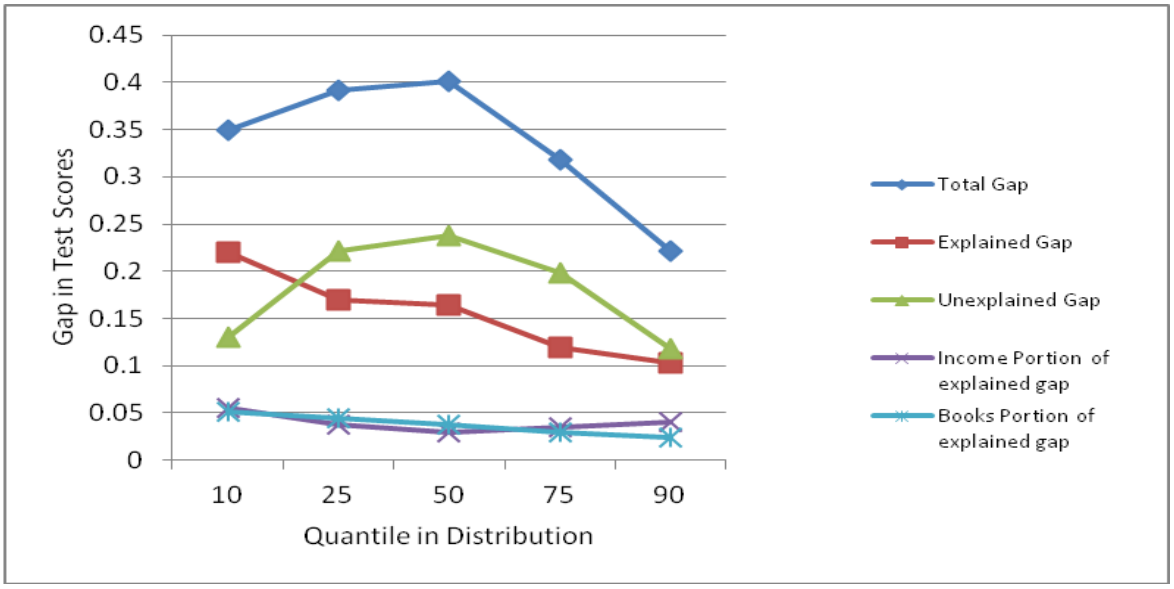

Figure 4b, Quantile Decompositions, Groups 4 and 2, Reading

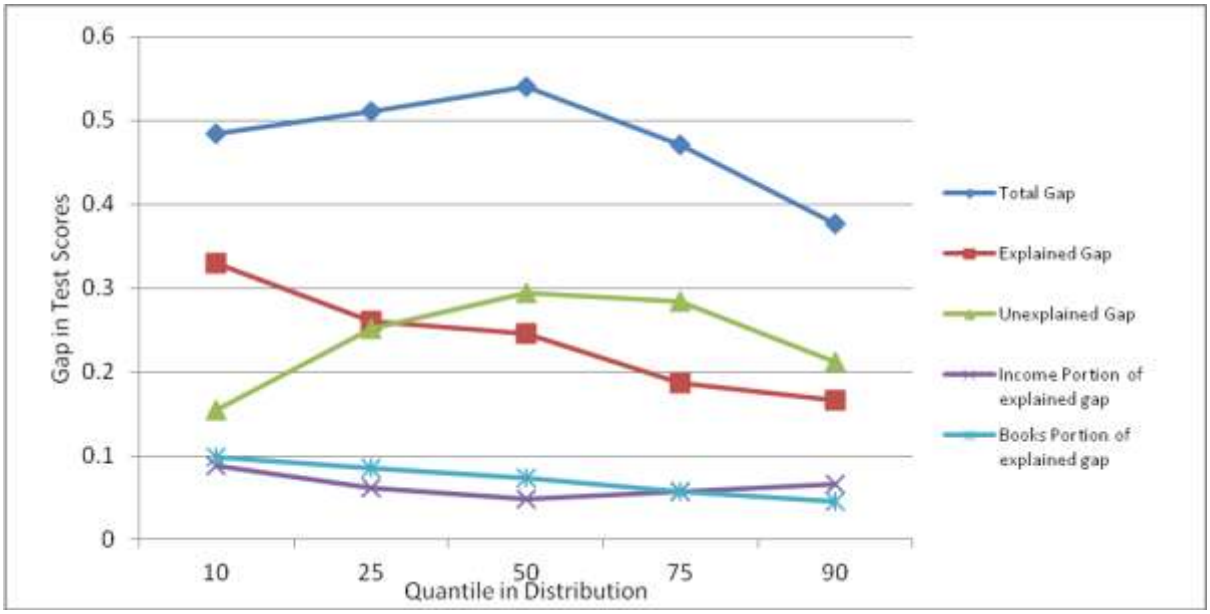

Figure 4c: Quantile Decompositions, Groups 4 and 1, Reading

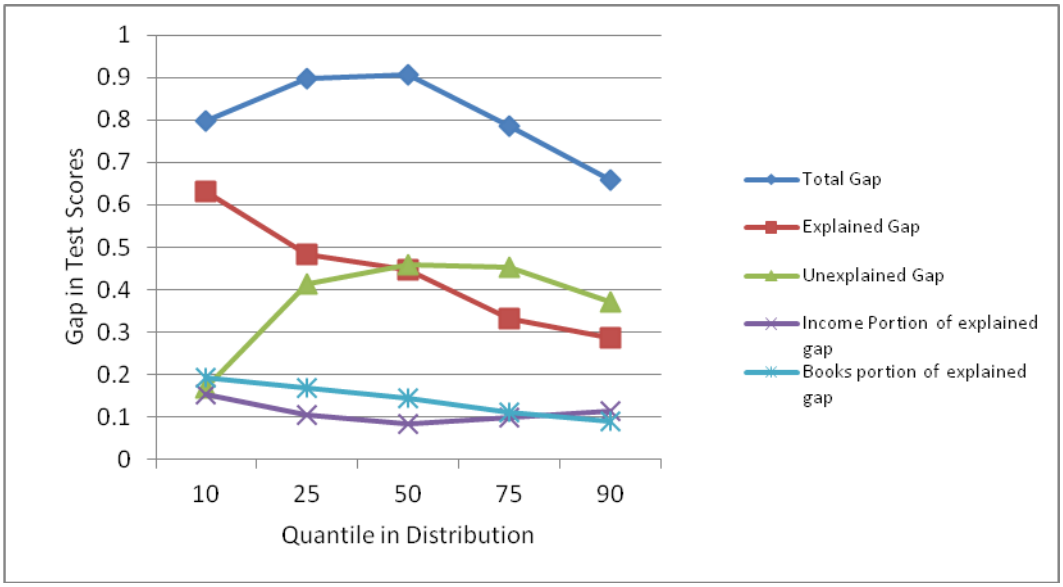

Figure 4d, Quantile Decompositions, Groups 3 and 2, Reading

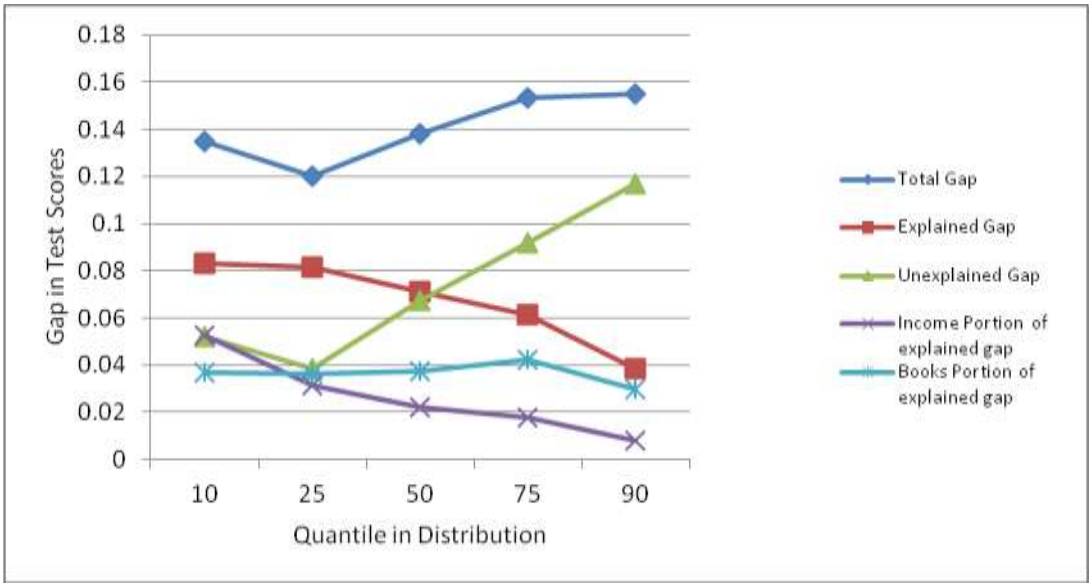

Figure 4e: Quantile Decompositions, Groups 3 and 1, Reading

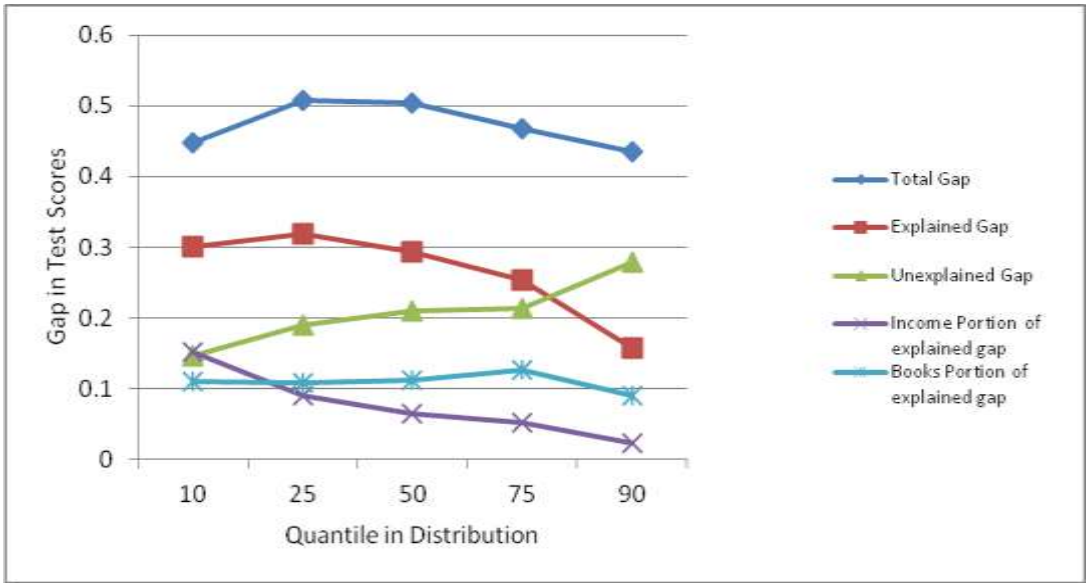

Figure 4f: Quantile Decompositions, Groups 2 and 1, Reading

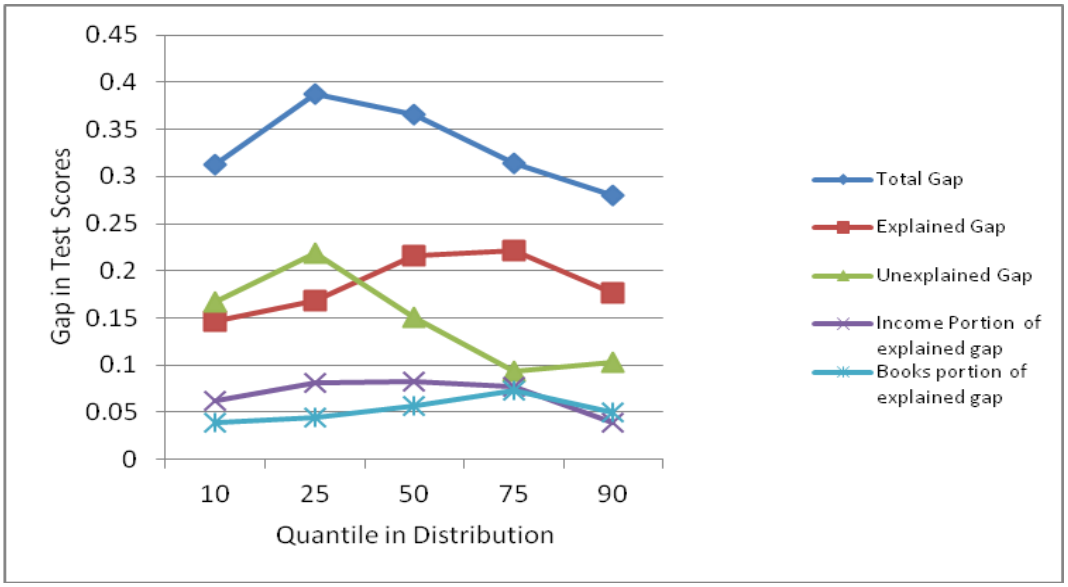

Table A1: Definition of Variables

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 A2: Detailed Quantile Decomposition, Maths, Groups 4 and 3

	q(25)		q(50)		q(75)	
Predicted (4)	-0.940***		-0.351***		0.2507***	
Predicted (3)	-1.204***		-0.575***		0.0314	
Gap	0.264***		0.224***		0.2193***	
	Explained	Unexp.	Explained	Unexp.	Explained	Unexp.
Age	0.0213	-0.0117	0.0291**	0.0522	0.0228*	-0.3199
Birthweight	0.0005	-0.0925	0.0004	-0.3992	0.0004	-0.1159
Early birth	0.0010	-0.0017	0.0022	-0.0084	0.0010	-0.0021
Smoker	0.0134	-0.0385	0.0064	-0.0036	0.0085	0.0051
Preg Smoker	0.0016	-0.0037	0.0040	-0.0178	0.0021	0.0006
Preg Drinker	0.0006	0.0043	0.0007	0.0059	0.0003	0.0056
Breastfed	0.0318**	0.1375**	0.0283**	0.0935*	0.0167	0.0541
Illness	0.0010	0.0097	0.0011	-0.0078	0.0007	-0.0067
Log Eqinc	0.0472***	0.6833	0.0441***	-0.1596	0.0347**	-0.2586
Mum healthy	0.0009	0.0787	0.0006	-0.0095	0.0002	-0.0431
Trauma	-0.0008	-0.0806	-0.0007	-0.0797	-0.0011	-0.0355
Books	0.0178*	0.0634	0.0293***	0.1165	0.0193**	0.1234
Local 1	0.0051	0.3504	0.0036	-0.0043	-0.0018	-0.0548
Local 2	0.0028	0.0093	0.0008	0.1886	0.0024	0.0573
Working	-0.0179*	-0.0693	-0.0141*	-0.0612	-0.0264***	-0.1841***
Size class	-0.0056	-0.7369***	-0.0022	-0.2209	-0.0017	-0.0059
Par/teacher	0.0003	-0.0126	0.0000	-0.0355	0.0018	0.0822
Engage	0.0004	0.1898	0.0011	0.1528	0.0016	0.0282
T experience	-0.0015	0.0390	-0.0008	-0.0361	-0.0004	-0.0941*
Quality	0.0034	0.0155	0.0025	-0.0102	-0.0001	-0.0041
School size	-0.0010	-0.0062	-0.0030	-0.0031	-0.0046	0.0092
Young sib	0.0023	-0.0436	-0.0028	-0.0931*	-0.0013	-0.0007
Old sib	0.0006	0.0452	-0.0000	-0.0150	-0.0000	-0.0048
Partner	-0.0014	-0.2412	0.0005	0.0353	-0.0012	-0.1946
Constant		-0.1473		0.6130		1.1046
Total	0.124***	0.140***	0.131***	0.0927*	0.0737***	0.1455***

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

Table A3: Detailed Quantile Decomposition, Maths, Groups 4 and 2

	q(25)		q(50)		q(75)	
Predicted (4)	-0.9403***		-0.3513***		0.2507***	
Predicted (2)	-1.2987***		-0.6594***		-0.0731**	
Gap	0.3584***		0.3081***		0.3238***	
	Explained	Unexp.	Explained	Unexp.	Explained	Unexp.
Age	0.0195	-0.2934	0.0266**	-0.1522	0.0208*	-0.6752
Birthweight	0.0014	0.0960	0.0011	-0.0679	0.0012	-0.0713
Early birth	0.0011	0.0015	0.0024	-0.0035	0.0011	0.0015
Smoker	0.0186	-0.0468	0.0088	-0.0181	0.0118	-0.0217
Preg Smoker	0.0025	-0.0012	0.0065	-0.0120	0.0035	-0.0084
Preg Drinker	0.0007	0.0051	0.0009	0.0062	0.0004	-0.0004
Breastfed	0.0622**	0.0736*	0.0555**	0.0546	0.0327	0.0112
Illness	0.0012	0.0170	0.0013	0.0087	0.0008	-0.0013
Log Eqinc	0.0765***	0.5354	0.0714***	0.1888	0.0562**	-1.0793
Mum healthy	0.0018	0.0683	0.0013	0.1522	0.0003	0.1299
Trauma	-0.0041	-0.0294	-0.0034	-0.0186	-0.0056	-0.0077
Books	0.0343*	-0.0376	0.0567***	0.2680	0.0374**	0.2229
Local 1	0.0066	0.1772	0.0047	-0.0161	-0.0024	-0.3169
Local 2	0.0027	-0.0017	0.0007	0.0868	0.0023	0.0133
Working	-0.0372*	-0.0729	-0.0293*	-0.0566	-0.0550***	-0.0930*
Size class	-0.0073*	-0.5746***	-0.0029	-0.2749	-0.0023	-0.2880
Par/teacher	0.0000	-0.0068	0.0000	-0.0347	0.0002	0.2440*
Engage	0.0004	0.0245	0.0013	0.1020	0.0020	0.1817
T experience	-0.0040	0.0115	-0.0021	-0.0074	-0.0011	-0.0116
Quality	0.0043	0.0291	0.0032	-0.0054	-0.0002	0.0217
School size	0.0004	-0.0044	0.0013	0.0048	0.0020	0.0117
Young sib	0.0054	-0.0183	-0.0066	-0.0619	-0.0031	-0.0441
Old sib	-0.0047	0.0304	0.0000	0.0175	0.0002	0.0340
Partner	-0.0013	-0.1730	0.0005	-0.0331	-0.0012	-0.0032
Constant		0.3675		-0.0193		1.9718*
Total	0.1813***	0.1771***	0.2000***	0.1081*	0.1020**	0.2218***

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

Table A4: Detailed Quantile Decomposition, Maths, Groups 4 and 1

	q(25)		q(50)		q(75)	
Predicted (4)	-0.9403***		-0.3513***		0.2507***	
Predicted (1)	-1.7057***		-1.0690***		-0.4751***	
Gap	0.7653***		0.7176***		0.7257***	
	Explained	Unexp.	Explained	Unexp.	Explained	Unexp.
Age	0.0311	0.2089	0.0425**	0.0448	0.0333*	-0.0302
Birthweight	0.0084	0.0017	0.0067	0.3188	0.0070	0.3222
Early birth	0.0034	-0.0025	0.0075	-0.0069	0.0035	0.0002
Smoker	0.0547	-0.1144*	0.0259	-0.0462	0.0347	-0.0505
Preg Smoker	0.0079	0.0540	0.0202	0.0155	0.0108	0.0107
Preg Drinker	-0.0007	-0.0014	-0.0009	0.0001	-0.0004	-0.0012
Breastfed	0.0969**	0.0433	0.0864**	0.0258	0.0509	-0.0075
Illness	0.0050	0.0079	0.0054	-0.0068	0.0034	0.0025
Log Eqinc	0.1316***	0.1998	0.1229***	0.8419	0.0967**	0.4619
Mum healthy	0.0098	0.0529	0.0071	0.0802	0.0018	0.0168
Trauma	-0.0021	-0.0844	-0.0017	-0.0160	-0.0029	-0.0511
Books	0.0675*	-0.1927	0.1114***	0.2795	0.0734**	0.1792
Local 1	0.0246	-0.0951	0.0174	0.0458	-0.0089	-0.3051
Local 2	0.0045	-0.1542	0.0012	-0.0633	0.0038	0.0093
Working	-0.0588*	-0.0587	-0.0463*	-0.0755*	-0.0869***	-0.0715*
Size class	-0.0300**	-0.5553**	-0.0119	-0.3780*	-0.0094	-0.2219
Par/teacher	0.0012	-0.0975	0.0001	-0.0984	0.0070	0.1398
Engage	0.0003	-0.0913	0.0008	-0.0351	0.0012	0.0596
T experience	0.0018	0.0522	0.0010	0.0089	0.0005	0.0024
Quality	0.0064	0.0003	0.0048	-0.0065	-0.0002	-0.0155
School size	-0.0002	0.0049	-0.0006	0.0038	-0.0010	0.0098
Young sib	0.0073	0.0308	-0.0091	-0.0116	-0.0042	-0.0128
Old sib	-0.0123	0.1035	0.0001	0.0335	0.0004	0.0893
Partner	-0.0066	-0.1940	0.0026	-0.0116	-0.0058	-0.0812
Constant		1.2952		-0.6185		0.0618
Total	0.3516***	0.4137***	0.3934***	0.3242***	0.2086***	0.5171***

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

Table A5: Detailed Quantile Decomposition, Maths, Groups 3 and 2

	q(25)		q(50)		q(75)	
Predicted (3)	-1.2042***		-0.5752***		0.0314	
Predicted (2)	-1.2987***		-0.6594***		-0.0731**	
Gap	0.0945**		0.0842*		0.1045*	
	Explained	Unexp.	Explained	Unexp.	Explained	Unexp.
Age	-0.0018	-0.2817	-0.0023	-0.2045	-0.0030	-0.3541
Birthweight	0.0013	0.1882	0.0021	0.3299	0.0012	0.0442
Early birth	0.0000	0.0033	-0.0001	0.0053	0.0000	0.0037
Smoker	-0.0013	-0.0019	0.0018	-0.0140	0.0041	-0.0277
Preg Smoker	-0.0002	0.0038	-0.0033	0.0117	0.0016	-0.0092
Preg Drinker	-0.0000	0.0010	-0.0001	0.0006	-0.0002	-0.0058
Breastfed	-0.0065	-0.0270	0.0020	-0.0138	0.0015	-0.0283
Illness	0.0004	0.0071	0.0000	0.0167	-0.0000	0.0055
Log Eqinc	0.0199*	-0.1385	0.0295**	0.3463	0.0250*	-0.8243
Mum healthy	0.0003	-0.0098	0.0008	0.1616	0.0005	0.1727
Trauma	0.0035	0.0444	0.0041	0.0544	-0.0015	0.0248
Books	0.0138*	-0.0981	0.0223**	0.1566	0.0126	0.1049
Local 1	-0.0004	-0.1713	0.0011	-0.0118	-0.0002	-0.2624
Local 2	-0.0001	-0.0109	-0.0003	-0.1015	-0.0002	-0.0439
Working	-0.0061	-0.0168	-0.0035	-0.0071	0.0067	0.0559
Size class	0.0015	0.1591	0.0003	-0.0550	-0.0005	-0.2821
Par/teacher	-0.0004	0.0059	-0.0004	0.0012	-0.0007	0.1609
Engage	-0.0004	-0.1649	-0.0001	-0.0504	0.0003	0.1536
T experience	-0.0014	-0.0285	-0.0024	0.0297	-0.0033	0.0852
Quality	0.0004	0.0141	0.0009	0.0045	0.0001	0.0257
School size	0.0040	-0.0008	0.0056	0.0066	0.0027	0.0064
Young sib	0.0085*	0.0199	0.0076	0.0197	-0.0017	-0.0434
Old sib	0.0001	-0.0202	-0.0018	0.0344	-0.0004	0.0394
Partner	-0.0001	0.0684	0.0000	-0.0685	-0.0001	0.1915
Constant		0.5147		-0.6323		0.8672
Total	0.0350*	0.0595	0.0640***	0.0203	0.0444*	0.0601

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

Table A6: Detailed Quantile Decomposition, Maths, Groups 3 and 1

	q(25)		q(50)		q(75)	
Predicted (3)	-1.2042***		-0.5752***		0.0314	
Predicted (1)	-1.7057***		-1.0690***		-0.4751***	
Gap	0.5015***		0.4938***		0.5065***	
	Explained	Unexp.	Explained	Unexp.	Explained	Unexp.
Age	0.0101	0.2204	0.0124*	-0.0063	0.0166*	0.2836
Birthweight	0.0106*	0.0914	0.0180**	0.7063**	0.0100	0.4347
Early birth	0.0010	0.0006	-0.0019	0.0087	0.0007	0.0042
Smoker	-0.0103	-0.0244	0.0147	-0.0378	0.0330	-0.0625
Preg Smoker	-0.0015	0.0656	-0.0212	0.0708	0.0100	0.0088
Preg Drinker	0.0003	-0.0073	0.0007	-0.0080	0.0015	-0.0089
Breastfed	-0.0139	-0.0152	0.0043	-0.0140	0.0031	-0.0305
Illness	0.0086	-0.0065	0.0006	0.0047	-0.0005	0.0125
Log Eqinc	0.0575*	-0.4566	0.0851**	0.9953	0.0722*	0.7103
Mum healthy	0.0032	-0.0200	0.0072	0.0890	0.0048	0.0567
Trauma	0.0014	-0.0064	0.0016	0.0611	-0.0006	-0.0168
Books	0.0414*	-0.2478	0.0668***	0.1783	0.0379*	0.0719
Local 1	-0.0050	-0.4210*	0.0141	0.0498	-0.0032	-0.2541
Local 2	0.0019	-0.1637	0.0044	-0.2558	0.0026	-0.0492
Working	-0.0128	-0.0175	-0.0074	-0.0392	0.0141	0.0379
Size class	0.0207**	0.1365	0.0039	-0.1707	-0.0073	-0.2164
Par/teacher	0.0013	-0.0854	0.0013	-0.0641	0.0023	0.0604
Engage	0.0005	-0.2817	0.0002	-0.1884	-0.0004	0.0313
T experience	0.0019	0.0146	0.0031	0.0437	0.0044	0.0931
Quality	0.0014	-0.0136	0.0033	0.0026	0.0003	-0.0119
School size	0.0022	0.0098	0.0031	0.0062	0.0015	0.0027
Young sib	0.0138*	0.0656	0.0124	0.0628	-0.0028	-0.0122
Old sib	0.0003	0.0450	-0.0043	0.0530	-0.0010	0.0955
Partner	0.0120	0.0299	-0.0005	-0.0443	0.0093	0.0995
Constant		1.4424		-1.2315		-1.0428
Total	0.1466***	0.3548***	0.2218***	0.2720***	0.2086***	0.2978***

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

Table A7: Detailed Quantile Decomposition, Maths, Groups 2 and 1

	q(25)		q(50)		q(75)	
Predicted (2)	-1.2987***		-0.6594***		-0.0731**	
Predicted (1)	-1.7057***		-1.0690***		-0.4751***	
Gap	0.4070***		0.4095***		0.4019***	
	Explained	Unexp.	Explained	Unexp.	Explained	Unexp.
Age	0.0182**	0.4957	0.0193**	0.1936	0.0275***	0.6298
Birthweight	0.0044	-0.0918	0.0073	0.3850	0.0076	0.3917
Early birth	0.0035	-0.0053	0.0023	-0.0007	0.0035	-0.0024
Smoker	-0.0109	-0.0206	-0.0011	-0.0098	0.0011	-0.0071
Preg Smoker	0.0038	0.0567	-0.0023	0.0435	-0.0039	0.0304
Preg Drinker	0.0008	-0.0087	0.0010	-0.0089	-0.0009	-0.0006
Breastfed	0.0039	0.0005	0.0081	-0.0060	0.0135	-0.0141
Illness	0.0114*	-0.0168	0.0080*	-0.0194	0.0020	0.0044
Log Eqinc	0.0412**	-0.3217	0.0466**	0.6580	0.0687***	1.5131
Mum healthy	0.0035	-0.0108	-0.0042	-0.0620	-0.0071	-0.1046
Trauma	0.0003	-0.0533	0.0006	0.0036	0.0023	-0.0429
Books	0.0365***	-0.1586	0.0303***	0.0359	0.0157	-0.0234
Local 1	0.0065	-0.2608	0.0138	0.0608	0.0140	-0.0087
Local 2	0.0018	-0.1525	0.0024	-0.1520	0.0018	-0.0043
Working	-0.0023	-0.0051	-0.0020	-0.0339	-0.0073	-0.0032
Size class	0.0101	-0.0135	0.0068	-0.1189	0.0093	0.0496
Par/teacher	0.0015	-0.0910	0.0016	-0.0653	-0.0041	-0.0933
Engage	-0.0000	-0.1159	0.0000	-0.1377	0.0002	-0.1231
T experience	0.0051	0.0413	0.0036	0.0158	0.0024	0.0132
Quality	-0.0001	-0.0267	0.0020	-0.0015	-0.0017	-0.0356
School size	-0.0021	0.0108	-0.0004	-0.0025	0.0008	-0.0058
Young sib	0.0036	0.0475	0.0031	0.0448	0.0028	0.0274
Old sib	-0.0029	0.0684	0.0028	0.0133	0.0055	0.0501
Partner	0.0072	-0.0335	0.0044	0.0192	-0.0044	-0.0782
Constant		0.9277		-0.5992		-1.9100
Total	0.1450***	0.2619***	0.1538***	0.2557***	0.1495***	0.2525***

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

Table A8: Detailed Quantile Decomposition, Reading, Groups 4 and 3

	q(25)		q(50)		q(75)	
Predicted (4)	-0.0704*		0.6142***		1.1880***	
Predicted (3)	-0.4614***		0.2128***		0.8705***	
Gap	0.3910***		0.4014***		0.3175***	
	Explained	Unexp.	Explained	Unexp.	Explained	Unexp.
Age	0.0512***	-0.4336	0.0685***	0.3510	0.0436***	-0.4005
Birthweight	0.0008	-0.1874	0.0007	0.0701	0.0003	-0.1580
Early birth	0.0005	-0.0011	0.0011	-0.0061	-0.0002	-0.0033
Smoker	0.0058	-0.0264	0.0106	-0.0247	0.0020	0.0168
Preg Smoker	-0.0008	0.0106	-0.0023	0.0117	0.0075	-0.0167
Preg Drinker	0.0003	-0.0022	-0.0003	-0.0060	0.0004	-0.0001
Breastfed	0.0288*	0.1044	0.0250*	0.0745	0.0185*	0.0989*
Illness	0.0015	0.0136	0.0019	-0.0089	0.0013	0.0066
Log Eqinc	0.0377**	-0.5913	0.0300*	-0.2578	0.0353***	0.3088
Mum healthy	0.0004	-0.1051	-0.0005	-0.3680*	-0.0006	-0.1819
Trauma	-0.0006	-0.0252	-0.0004	-0.0555	-0.0003	-0.0490
Books	0.0443***	0.1147	0.0383***	-0.0339	0.0293***	-0.3420
Local 1	-0.0021	-0.0537	-0.0012	-0.2145	-0.0032	-0.1034
Local 2	0.0040	-0.0134	0.0012	0.1684	-0.0009	0.2226
Working	-0.0086	-0.0547	-0.0129	-0.0531	-0.0094	-0.0173
Size class	0.0010	-0.0427	-0.0027	-0.3468	-0.0029	-0.3137
Par/teacher	0.0029	0.1241	0.0001	-0.0599	0.0006	0.0631
Engage	0.0010	0.0336	0.0006	0.0749	0.0000	-0.0685
T experience	-0.0002	-0.0395	0.0005	-0.0659	-0.0001	-0.0130
Quality	0.0036	0.0173	0.0066	0.0507	0.0017	-0.0191
School size	-0.0042	0.0139	-0.0043	0.0168	-0.0033	0.0033
Young sib	0.0029	-0.0203	0.0027	0.0453	0.0014	0.0512
Old sib	-0.0002	0.0650	-0.0011	0.0091	-0.0006	0.0794
Partner	-0.0003	0.0248	0.0022	0.1401	-0.0010	-0.0179
Constant		1.2958		0.7256		1.0516
Total	0.1699***	0.2211***	0.1644***	0.2370***	0.1196***	0.1979***

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

Table A9: Detailed Quantile Decomposition, Reading, Groups 4 and 2

	q(25)		q(50)		q(75)	
Predicted (4)	-0.0704*		0.6142***		1.1880***	
Predicted (2)	-0.5813***		0.0744**		0.7172***	
Gap	0.5109***		0.5398***		0.4708***	
	Explained	Unexp.	Explained	Unexp.	Explained	Unexp.
Age	0.0468***	0.0678	0.0627***	0.2985	0.0399***	-0.7599*
Birthweight	0.0024	0.3501	0.0020	0.3198	0.0008	-0.1999
Early birth	0.0005	0.0052	0.0012	0.0017	-0.0002	0.0043
Smoker	0.0080	0.0034	0.0146	-0.0239	0.0027	0.0077
Preg Smoker	-0.0014	-0.0088	-0.0037	0.0033	0.0123	-0.0339
Preg Drinker	0.0004	0.0029	-0.0003	-0.0007	0.0004	0.0020
Breastfed	0.0565*	0.0616	0.0490*	0.0234	0.0362*	-0.0051
Illness	0.0018	-0.0029	0.0022	-0.0043	0.0016	-0.0102
Log Eqinc	0.0611**	-1.4123	0.0487*	-1.8155*	0.0573***	-1.3640
Mum healthy	0.0008	0.1926	-0.0010	0.0020	-0.0013	-0.0110
Trauma	-0.0029	0.0073	-0.0019	0.0139	-0.0013	-0.0115
Books	0.0857***	0.4263*	0.0739***	0.1599	0.0567***	-0.1999
Local 1	-0.0027	-0.0015	-0.0015	-0.0701	-0.0042	-0.1333
Local 2	0.0039	0.0644	0.0012	0.1300	-0.0008	0.1514
Working	-0.0179	-0.0152	-0.0269	-0.0380	-0.0196	-0.0100
Size class	0.0014	0.0784	-0.0035	-0.2954	-0.0038	-0.1672
Par/teacher	0.0003	0.1713	0.0000	-0.0618	0.0001	0.0433
Engage	0.0012	0.1063	0.0007	0.0527	0.0000	0.0596
T experience	-0.0006	-0.0318	0.0015	-0.0735	-0.0001	-0.0363
Quality	0.0045	0.0126	0.0082*	0.0592*	0.0021	-0.0060
School size	0.0018	0.0113	0.0019	0.0051	0.0015	0.0084
Young sib	0.0069	-0.0207	0.0064	-0.0107	0.0033	-0.0274
Old sib	0.0014	0.0167	0.0079	0.0192	0.0042	0.1063*
Partner	-0.0002	0.0609	0.0021	0.2293	-0.0009	0.1492
Constant		0.1053		1.3704		2.7274**
Total	0.2597***	0.2512***	0.2454***	0.2944***	0.1870***	0.2839***

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

Table A10: Detailed Quantile Decomposition, Reading, Groups 4 and 1

	q(25)		q(50)		q(75)	
Predicted (4)	-0.0704*		0.6142***		1.1880***	
Predicted (1)	-0.9691***		-0.2914***		0.4030***	
Gap	0.8987***		0.9056***		0.7849***	
	Explained	Unexp.	Explained	Unexp.	Explained	Unexp.
Age	0.0748***	0.1800	0.1001***	0.7247*	0.0637***	-0.1967
Birthweight	0.0137*	0.2124	0.0118*	0.0751	0.0045	-0.1640
Early birth	0.0016	0.0016	0.0038	-0.0025	-0.0006	0.0046
Smoker	0.0236	-0.0432	0.0431	-0.0446	0.0081	0.0406
Preg Smoker	-0.0043	0.0346	-0.0116	0.0264	0.0384	-0.0754
Preg Drinker	-0.0004	-0.0049	0.0003	-0.0039	-0.0004	-0.0012
Breastfed	0.0880*	0.0430	0.0763*	0.0155	0.0563*	-0.0471
Illness	0.0075	0.0182	0.0091	-0.0203	0.0066	-0.0140
Log Eqinc	0.1052**	0.5423	0.0838*	0.0411	0.0986***	1.1967
Mum healthy	0.0041	-0.0947	-0.0056	-0.1154	-0.0067	-0.1072
Trauma	-0.0015	-0.0453	-0.0010	-0.0569	-0.0006	-0.0564
Books	0.1683***	0.0548	0.1452***	0.0762	0.1114***	-0.1869
Local 1	-0.0100	-0.4697*	-0.0056	-0.2118	-0.0157	-0.1807
Local 2	0.0065	-0.1849	0.0020	-0.1675	-0.0014	-0.1224
Working	-0.0283	-0.0457	-0.0425	-0.0509	-0.0310	-0.1142**
Size class	0.0057	0.1740	-0.0144	-0.1213	-0.0155	-0.2580
Par/teacher	0.0112*	0.1754	0.0003	-0.0078	0.0023	0.1493
Engage	0.0007	-0.2529	0.0004	-0.2448	0.0000	-0.1536
T experience	0.0003	0.0255	-0.0007	-0.0328	0.0001	0.0158
Quality	0.0067	0.0390	0.0123*	0.0244	0.0032	-0.0037
School size	-0.0009	0.0182	-0.0009	0.0257*	-0.0007	0.0095
Young sib	0.0095	0.0033	0.0087	0.0881*	0.0045	0.0587
Old sib	0.0037	0.1728*	0.0209*	0.1041	0.0111	0.2005**
Partner	-0.0012	0.0820	0.0105	-0.0182	-0.0047	-0.2345
Constant		-0.2216		0.3565		0.6935
Total	0.4843***	0.4143***	0.4462***	0.4593***	0.3316***	0.4533***

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

Table A11: Detailed Quantile Decomposition, Reading, Groups 3 and 2

	q(25)		q(50)		q(75)	
Predicted (3)	-0.4614***		0.2128***		0.8705***	
Predicted (2)	-0.5813***		0.0744**		0.7172***	
Gap	0.1199**		0.1384***		0.1533***	
	Explained	Unexp.	Explained	Unexp.	Explained	Unexp.
Age	-0.0058	0.5029	-0.0046	-0.0537	-0.0051	-0.3581
Birthweight	0.0022	0.5368*	0.0011	0.2500	0.0011	-0.0425
Early birth	0.0000	0.0064	-0.0001	0.0081	-0.0001	0.0078
Smoker	-0.0022	0.0343	-0.0001	0.0050	0.0036	-0.0119
Preg Smoker	0.0029	-0.0229	0.0024	-0.0122	-0.0007	-0.0117
Preg Drinker	0.0002	0.0050	0.0002	0.0050	0.0001	0.0021
Breastfed	-0.0004	-0.0148	0.0040	-0.0311	-0.0089	-0.0775*
Illness	0.0006	-0.0168	0.0001	0.0048	0.0004	-0.0170
Log Eqinc	0.0315**	-0.8291	0.0222*	-1.5612	0.0177	-1.6686
Mum healthy	0.0012	0.2968	0.0023	0.3672*	0.0008	0.1695
Trauma	-0.0002	0.0303	0.0032	0.0647	0.0031	0.0334
Books	0.0363***	0.3166	0.0372***	0.1923	0.0423***	0.1271
Local 1	-0.0003	0.0519	0.0008	0.1432	-0.0004	-0.0305
Local 2	-0.0002	0.0777	-0.0003	-0.0381	-0.0003	-0.0708
Working	0.0012	0.0290	-0.0038	0.0049	-0.0069	0.0039
Size class	0.0005	0.1209	0.0007	0.0498	0.0005	0.1451
Par/teacher	-0.0013	0.0459	-0.0007	-0.0013	0.0001	-0.0205
Engage	0.0001	0.0729	-0.0001	-0.0220	0.0002	0.1280
T experience	-0.0015	0.0088	-0.0009	-0.0058	-0.0005	-0.0229
Quality	0.0004	-0.0042	0.0002	0.0100	0.0010	0.0126
School size	0.0002	0.0032	-0.0009	-0.0047	0.0033	0.0065
Young sib	0.0065	-0.0029	-0.0019	-0.0504	-0.0044	-0.0723
Old sib	0.0094	-0.0561	0.0101*	0.0091	0.0143*	0.0174
Partner	0.0000	0.0361	0.0000	0.0891	0.0000	0.1670
Constant		-1.1905		0.6448		1.6758
Total	0.0814***	0.0385	0.0710***	0.0674	0.0613**	0.0920*

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

Table A12: Detailed Quantile Decomposition, Reading, Groups 3 and 1

	q(25)		q(50)		q(75)	
Predicted (3)	-0.4614***		0.2128***		0.8705***	
Predicted (1)	-0.9691***		-0.2914***		0.4030***	
Gap	0.5077***		0.5042***		0.4674***	
	Explained	Unexp.	Explained	Unexp.	Explained	Unexp.
Age	0.0319**	0.6053	0.0250**	0.3803	0.0277**	0.1962
Birthweight	0.0184**	0.3943	0.0090	0.0071	0.0089	-0.0107
Early birth	0.0002	0.0037	-0.0025	0.0089	-0.0033	0.0107
Smoker	-0.0176	0.0187	-0.0006	0.0132	0.0285	0.0014
Preg Smoker	0.0188	0.0018	0.0153	-0.0100	-0.0044	-0.0234
Preg Drinker	-0.0015	-0.0018	-0.0017	0.0044	-0.0008	-0.0010
Breastfed	-0.0008	-0.0014	0.0085	-0.0162	-0.0190	-0.0892**
Illness	0.0124*	-0.0019	0.0030	-0.0071	0.0084	-0.0238
Log Eqinc	0.0908**	1.1103	0.0639*	0.2888	0.0511	0.9001
Mum healthy	0.0114	0.0027	0.0218*	0.2257	0.0072	0.0614
Trauma	-0.0001	-0.0209	0.0012	-0.0033	0.0012	-0.0090
Books	0.1090***	-0.0448	0.1114***	0.1057	0.1268***	0.1103
Local 1	-0.0042	-0.4198	0.0105	-0.0123	-0.0052	-0.0846
Local 2	0.0022	-0.1712	0.0043	-0.3394*	0.0041	-0.3496
Working	0.0025	-0.0132	-0.0081	-0.0193	-0.0145	-0.1039*
Size class	0.0072	0.2141	0.0095	0.2042	0.0066	0.0365
Par/teacher	0.0040	0.0556	0.0023	0.0500	-0.0004	0.0884
Engage	-0.0002	-0.2866	0.0001	-0.3199*	-0.0002	-0.0849
T experience	0.0019	0.0635	0.0012	0.0307	0.0006	0.0284
Quality	0.0013	0.0235	0.0005	-0.0210	0.0035	0.0135
School size	0.0001	0.0075	-0.0005	0.0127	0.0018	0.0070
Young sib	0.0106	0.0196	-0.0031	0.0520	-0.0071	0.0178
Old sib	0.0229*	0.0887	0.0246*	0.0924	0.0349**	0.0978
Partner	-0.0027	0.0590	-0.0018	-0.1483	-0.0024	-0.2179
Constant		-1.5174		-0.3691	0.2541***	-0.3581
Total	0.3185***	0.1892**	0.2939***	0.2103***	0.2541***	0.2134**

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

Table A13: Detailed Quantile Decomposition, Reading, Groups 2 and 1

	q(25)		q(50)		q(75)	
Predicted (2)	-0.5813***		0.0744**		0.7172***	
Predicted (1)	-0.9691***		-0.2914***		0.4030***	
Gap	0.3878***		0.3658***		0.3141***	
	Explained	Unexp.	Explained	Unexp.	Explained	Unexp.
Age	0.0264***	0.1137	0.0307***	0.4329	0.0408***	0.5462
Birthweight	0.0022	-0.1286	0.0014	-0.2364	0.0089	0.0307
Early birth	0.0052	-0.0077	0.0039	-0.0055	0.0029	-0.0031
Smoker	0.0190	-0.0500	0.0045	0.0032	0.0130	0.0252
Preg Smoker	-0.0148	0.0553	-0.0035	0.0187	-0.0193	0.0039
Preg Drinker	0.0005	-0.0091	0.0003	-0.0029	0.0000	-0.0041
Breastfed	0.0057	0.0072	0.0175*	0.0019	0.0223*	-0.0441
Illness	0.0044	0.0224	0.0050	-0.0140	0.0005	0.0008
Log Eqinc	0.0809***	1.9178	0.0825***	1.8093	0.0769***	2.5252*
Mum healthy	-0.0094	-0.2746	-0.0047	-0.1173	-0.0047	-0.0969
Trauma	0.0018	-0.0530	0.0017	-0.0716	-0.0000	-0.0442
Books	0.0438***	-0.3326*	0.0567***	-0.0691	0.0729***	-0.0052
Local 1	-0.0072	-0.4683*	0.0004	-0.1463	-0.0029	-0.0561
Local 2	0.0041	-0.2507	0.0037	-0.3004*	0.0029	-0.2772
Working	-0.0064	-0.0346	-0.0056	-0.0229	-0.0087	-0.1068*
Size class	-0.0002	0.1001	0.0060	0.1572	-0.0021	-0.1003
Par/teacher	0.0032	0.0118	0.0031	0.0513	0.0004	0.1080
Engage	0.0001	-0.3598*	0.0000	-0.2978*	0.0003	-0.2136
T experience	0.0029	0.0553	0.0024	0.0361	0.0025	0.0498
Quality	0.0013	0.0273	-0.0004	-0.0303	0.0015	0.0019
School size	0.0010	0.0032	-0.0011	0.0189*	0.0006	-0.0016
Young sib	0.0044	0.0222	0.0033	0.0979*	0.0036	0.0836
Old sib	0.0049	0.1535*	0.0159**	0.0819	0.0233**	0.0777
Partner	-0.0054	0.0254	-0.0082	-0.2309*	-0.0145	-0.3729**
Constant		-0.3269		-1.0138		-2.0338
Total	0.1684***	0.2194***	0.2157***	0.1501**	0.2211***	0.0931

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