Inequality of opportunity at school in rural Bangladesh: to what extent are pupils' efforts shaped by family background?

Niaz Asadullah¹, Alain Trannoy², Sandy Tubeuf³, and Gaston Yalonetzky⁴

Work-in-progress draft. Please do not circulate without the authors' permission.

Abstract

The notion of inequality of opportunity draws a distinction between "legitimate" and "illegitimate" sources of differences in wellbeing outcomes. While legitimate differences can be attributed to effort and illegitimate differences to circumstances (beyond people's control), the cut between the two sources is not clear. Specifically, legitimate inequality may be undermined by the importance of the correlation between effort and circumstances (e.g. family background) as underlined by John Roemer. This paper focuses on evaluating the importance of the correlation between circumstances and effort when measuring inequalities of opportunity in education. The school experience and performance are particularly interesting because they strongly impact on future adult life. We use data from a unique survey on secondary school education in rural Bangladesh with two indicators of performance, 14 indicators of students' effort, and a large set of circumstances. We undertake an original decomposition method allowing us to explain both within and between schools decomposition. We find that circumstances contribute to half of the total variance in both Mathematics and English test scores. Circumstances matter more for between-school variations while efforts matter more for within-school variations for both subjects. The importance of the viewpoint adopted appears to matter particularly for the withinschool models. This result suggests that while the normative position on how to treat the correlation between circumstances and effort made little difference in health in France (10% according to Jusot et al. 2013), it does matter more in education, and this confirms the importance of social determinism at school.

JEL codes: 124.

Keywords: inequality of opportunity, inequality decomposition, education, circumstances, effort, variance, Bangladesh.

Acknowledgements: Gaston Yalonetzky gratefully acknowledges the financial support provided by LABEX AMSE for a research visit to the Aix-Marseille School of Economics. This research work was presented and received comments from colleagues at the HEAL seminar series (University of Lancaster), the 6th meeting of the ECINEQ in Luxemburg and the 14th LAGV workshop in Aix-en-Provence.

1. Introduction

The last few decades have witnessed the remarkable development of new quantitative methods for the measurement of inequality of opportunity in different social settings, based on variations in the understanding of this concept of distributional justice. All versions around the notion of inequality of opportunity distinguish between legitimate and illegitimate sources of inequality.

¹ University of Kuala Lumpur.

² Aix-Marseille School of Economics, EHESS and CNRS.

³ University of Leeds.

⁴ University of Leeds.

The former are due to factors for which the individual can be held responsible, whereas the latter stem from factors beyond the individual's control. In Roemer's terminology, these are efforts and circumstances, respectively (Roemer, 1998). The typical ethical prescription is that inequalities due to circumstances should be compensated for (principle of compensation); whereas those due to efforts, and hence legitimate, should be respected (principle of liberal reward; Fleurbaey, 2008).

In the case of education, previous research argues that effort at school is strongly impacted by circumstances as measured by family and social background. Students' school performance has been found to be highly correlated with parental income, education, cognitive abilities, and parents' own effort as measured by their aspirations for, communication and participation in their children's school matters (Haveman and Wolfe, 1995; Ermish and Francesconi, 2001; De Fraja et al, 2010). Pupils' effort at school is therefore significantly constrained by circumstances. Moreover, according to Roemer, the correlation between effort and circumstances must be measured and accounted for, so that only the "true" contribution of effort to differences in school performance is respected.

In this paper we aim to evaluate empirically the importance of the correlation between effort at school and circumstances, and measure the magnitude of inequality of opportunity in school performance when implementing the principle of natural reward.

The main challenge for an empirical evaluation is the availability of detailed data on circumstances and efforts (in addition to the welfare outcomes over which inequality is considered). In the case of circumstances, their limited availability has led to acknowledging that our measures of inequality of opportunity usually provide lower bounds (Ferreira and Gignoux, 2011). Meanwhile, data on efforts are much harder to come by. One strand of the literature has implicitly dealt with this dearth of information by focusing on concepts of inequality of opportunity in which efforts do not play an active part. Specifically the so-called ex-ante approach compares distributional standards (e.g. a mean) of opportunity sets belonging to different social groups (Van de Gaer, 1993; Ooghe et al., 2006; Lefranc et al., 2008, 2009; Checchi and Peragine, 2010; Fleurbaey and Peragine, 2013), and these sets are constructed by combining circumstance categories also called *types* by Roemer. The degree of inequality of opportunity is then attached to the degree of between-group (or between-type) inequality related to these opportunity sets (i.e. conditional distributions of the outcome of interest).

Another strand of the literature implicitly avoids the problem of lacking effort indicators by combining an ex-post approach to inequality of opportunity with a concept of relative effort whereby two people belonging to different types are deemed to have exerted the same effort if and only if they are in the same percentiles of their respective (and different) conditional distributions (Roemer, 1998; Checchi and Peragine, 2010; Fleurbaey and Peragine, 2013).

Recently though, Jusot et al. (2013) have proposed a method for the measurement of inequality of opportunity when effort variables are actually available. They measured inequality of opportunity in a health indicator in France, and used lifestyles as a proxy for effort in relation to health. Their method is based on a parametric decomposition of the variance (and the square coefficient of variation), and is capable of accommodating different views regarding the correlation between circumstances and efforts. Some authors (e.g. Roemer) claim that only inequality due to efforts unrelated to circumstances should be respected, whereas others (e.g. Barry) argue that all inequality due to efforts should not be compensated, and finally a third viewpoint is to respect parental effort in the application of the principle of natural reward whatever its consequences to the next generation (e.g. Swift). Jusot et al. (2013) used their method to assess the empirical relevance of three alternative views (Roemer's, Barry's, and Swift's) on the correlation between circumstances and efforts, and how they affect the division between legitimate and illegitimate inequality in health. Interestingly, they found little variation across the different views.

In the education literature, effort at school has basically been measured by homework and study time from the perspective of either the student (De Fraja et al. 2010; Kuehn and Landeras, 2013) or the teachers' (Eren and Henderson, 2011). Some studies have also considered variables combining study and effort together along with an identification variable. Stinebrickner and Stinebrickner (2008) focused on variables such as class attendance, sleeping, drinking, study efficiency, paid employment in combination with the availability of a roommate's computer game, while Metcalfe et al. (2010) identified effort using school performance according to television viewing and games of English football team.

In this paper, we use data from a unique survey undertaken among rural secondary students attending state schools and registered Madrasas (i.e. Islamic religious schools) in rural Bangladesh. The "Quality of Secondary School Madrasah Education in Bangladesh" (QSSMEB) dataset provides us with a rich set of 14 variables measuring students' effort for two study subjects: mathematics and English, along with detailed school specific data and family and social background characteristics. We measure and decompose inequality of opportunity following both Roemer's and Barry's views regarding the impact of the correlation between efforts and circumstances. We find that the degree of legitimate inequality due to students' effort is relatively small compared to the degree of illegitimate inequality and it is larger for English test scores (12% to 19%) than it is for Mathematics test scores (4% to 7%). More interestingly, we find that the correlation between effort and circumstances represents 40% of the contribution of effort to the total variance in both Mathematics and English test scores. This result suggests that, while the normative position on how to treat the correlation between circumstances and effort made little difference in health (10% according to Jusot et al. 2013), it does matter in education.

The rest of the paper proceeds as follows. Section two clarifies the two main views regarding how the correlation between efforts and circumstances should be handled when measuring inequality of opportunity. Section three explains the rural Bangladesh context, and describes key features of the dataset and the chosen wellbeing outcomes, effort indicators, and circumstances. Section fourth lays out the measurement method. Section five presents our results. Finally the paper concludes with some final remarks.

2. Views on the correlation between efforts and circumstances

As Jusot et al. (2013) explain, there are different views as to how the level of inequality attributable to differential effort should be respected, i.e. as a form of legitimate inequality. The problem is that, admittedly, in many situations, people's efforts may in themselves be related to more or less conducive circumstances, which by definition are beyond the individuals' control. Roemer (1998), for instance, argues that any effort due to circumstances should be compensated for. By contrast Barry (2005) praises the effort exerted, whether induced by circumstances or not. Therefore in this view the rewards of efforts (or lack thereof) should not be tampered with.

While Arneson (1990) championed the need to respect individual's responsibility for their preferences, he emphasised the importance of referring to an "age of consent" that acts as a threshold below which people cannot be held responsible for their effort. The concept of "age of consent" is particularly relevant in the case of education as education mainly happens in

childhood and teenage years (Roemer and Trannoy, 2015). While pupils are sometimes held responsible for their effort in doing homework (e.g. most empirical studies use study time to measure effort), the view on the cut between legitimate and illegitimate inequality in school performance varies. Lu et al. (2013) undertook an experiment to elicit views on salesmen and students' responsibility for their performance. Whereas salesmen were found responsible for their effort in undertaking homework by half of the respondents; most of the other factors related to their educational achievements were deemed independent from students' responsibility. The data we use do not include IQ, our measure of school performance and responsibility are based on observable reported variables, and therefore we cannot speak in terms of causality. If we interpret IQ as an unobserved circumstance (Feindstein, 2003), the way we carry the analysis is similar to the findings in Lu et al. (2013).

This literature suggests that we cannot treat equally similar efforts in handing over the homework in time, or paying attention in class, if they can be traced back to two different home environments. If wellbeing outcomes are positively associated with both circumstances and efforts, and the latter two are also positively associated between themselves, then Roemer's view yields a lower magnitude of legitimate inequality vis-à-vis Barry's.

3. The case of rural Bangladesh schools

a. Context

The educational system in rural Bangladesh is characterized by the prevalence of madrasahs, i.e. Islamic religious schools, operating alongside state schools (madrasah enrolment is about 14% at the primary level and 30% at the secondary level; Asadullah et al., 2009). While the madrasahs are thought to be run by motivated religious personnel and credited to offer a cheaper alternative to poorer people, they are also feared for the potential nurturing of militancy, but fundamentally criticized for offering education of poorer quality, thereby potentially perpetuating a poverty cycle (Asadullah, 2014).

However, in reality, there are two types of madrasahs in rural Bangladesh. Starting in the early 1980s, the government offered financial incentives to madrasahs in exchange for teaching the state curriculum and accepting female students. Most madrasahs took up the offer and became registered Aliya madrasahs. A minor unregulated sector remained, called Quami madrasah. Noticeably the initiative helped reduce the gender gap in female education. (Asadullah and Chaudhury, 2009).

Moreover, while both economic and religious factors affect parents' decision to send children to madrasahs, the former are more important (Asadullah et al., 2013). Considering that school quality in rural Bangladesh is generally poor across the board, there is no significant performance gap between state schools and registered madrasahs once school sorting is accounted for (Asadullah et al., 2007), however a trade-off between performance in English and in religious studies remain, whereby state schools perform better at English while madrasahs do better at Islamic knowledge (Asadullah, 2014).

In this paper our interest rather lies in quantifying the degree of legitimate and illegitimate sources of inequality in observed indicators of students' performance, under two different views prescribing how to draw the line between these two sources.

b. Data

The data come from a survey called "Quality of Secondary School Madrasah Education in Bangladesh" (QSSMEB) whose collection started in 2008 under the auspices of the World Bank in order to gauge the quality of education in registered madrasahs vis-à-vis state schools. The survey was co-designed by one of this paper's co-authors. Detailed information about the sampling procedure, scope and range of information provided by the survey can be found in Asadullah et al. (2009) and in Asadullah (2014). In this section we will focus on describing the data aspects pertaining to the choice of variables relevant to our inequality of opportunity assessment.

Educational outcomes

In every sampled union (a Bangladeshi sub-district larger than a village but smaller than subdistricts called upazilas) all secondary schools were surveyed. In each of them the surveyors administered four cognitive tests to 8th grade students. For our analysis we use two of these tests: first, a mathematics test using the 25 items of the Trends in Mathematics and Science Study (TIMSS);⁵ secondly, an English proficiency test with 20 items devised by the surveyors and based on the country's national curriculum (Asadullah, 2014). The distributions of scores for both tests are in figures 3.1 and 3.2. In line with the above mentioned evidence of poor quality of education in rural Bangladesh, most students failed to score above 50% of correct answers in each test.



Figure 3.1 Mathematics test scores

⁵ For further details see <u>www.timss.com</u>.

Figure 3.2 English test scores



Students' efforts

Based on the 1988 National Education Longitudinal Study (NELS; US Department of Education, 1988), both the mathematics and English teachers filled a subjective assessment of every sample student on seven aspects of students' behaviour in the classroom which we use as indicators of effort: (1) how often student performs below ability; (2) how often student submits incomplete homework; (3) how often student is absent; (4) how often student is tardy or lazy; (5) how often student is inattentive in class; (6) how often student is disinterested; (7) how often student makes noise (disruptive). For all questions the possible answers are: "Never", "Rarely", "Sometimes", "Somewhat", and "Always". The descriptive statistics are presented in tables 3.3 and 3.4, respectively for English and mathematics. The most common categories are "never", "rarely", and "sometimes".

For the inequality of opportunity assessment we have dichomotized the effort indicators by merging the "never" and "rarely" categories on one hand, and then merging "sometimes" with "somewhat" and "always" to generate the second binomial category.⁶

		0				
	Never	Rarely	Sometimes	Somewhat	Always	Sample
	(%)	(%)	(%)	(%)	(%)	sıze
How often bad result in respect to merit	29.59	39.47	23.02	6.56	1.35	8441
How often incomplete homework	22.98	46.21	24.41	5.89	0.51	8439
How often absent	39.14	37.16	19.40	3.70	0.60	8437
How often lazy or comes late	35.31	37.91	20.45	5.58	0.75	8440
How often inattentive	38.41	36.49	19.44	4.90	0.76	8437
How often disinterested in class work	67.99	21.94	8.00	1.74	0.33	8437
How often makes noise	21.12	78.88				8396

Table 3.3 Student's efforts indicators in English class

⁶ This way of proceeding spares us from the need to implement ordered multinomial models that do not rely on the proportional odds assumption, which is violated in our dataset whenever we model the effort indicators as a function of family circumstances. The latter is part of the procedure to measure inequality of opportunity following Roemer's approach (see Methodology section). Interestingly the original 1988 NELS only allowed binary responses; US Department of Education, 1988, p. 2-3.).

	Never (%)	Rarely (%)	Sometimes (%)	Somewhat (%)	Always (%)	Sample size
How often bad result in respect to merit	27.24	35.89	25.00	9.28	2.58	8219
How often incomplete homework	30.34	38.49	23.18	6.97	1.02	8223
How often absent	23.72	45.60	24.94	5.44	0.29	8211
How often lazy or comes late	42.05	35.87	18.46	3.19	0.44	8225
How often inattentive	38.23	37.57	19.33	4.26	0.62	8222
How often disinterested in class work	40.15	36.53	18.99	3.70	0.63	8224
How often makes noise	70.37	21.69	6.40	1.36	0.18	8694

Table 3.4 Student's efforts indicators in Mathematics class

Personal circumstances

For the assessment we consider student's age and gender. The respective descriptive statistics are in table 3.5. The sample is nearly 60% women. We use three age dummies: (1) Aged 13 years old; (2) aged 14 years old; (3) aged 15 years old or older. The omitted category is aged 12 years old or younger.

	Mean (%)	Standard deviation	Sample
Female	62.06	0.485	9021
Age=10	0.15		8847
Age=11	2.94		8847
Age=12	21.97		8847
Age=13	45.11		8847
Age=14	22.63		8847
Age=15	5.53		8847
Age=16	0.96		8847
Age=17	0.31		8847
Age=18	0.27		8847
Age=19	0.07		8847
Age=20	0.05		8847
Age=21	0.01		8847
Age=22	0.01		8847

Parental circumstances

For the measurement of parental circumstances we use dummies of educational attainment for both fathers and mothers. For each parent the dummies are: (1) if completed only primary education; (2) if did some secondary education; (3) if completed up to secondary education; (4) if did some tertiary education. The omitted category is incomplete primary education (or less). Households from religious minorities account for 7%. The educational and religious distributions are in table 3.6.

Table 3.6 Parental characteristics

	Mean (%)	Standard deviation	Sample size
Education			
Father complete primary	12.01	0.325	9021
Father incomplete secondary	16.79	0.374	9021
Father complete secondary	15.89	0.366	9021
Father some tertiary	17.75	0.382	9021
Mother complete primary	18.75	0.390	9021
Mother incomplete primary	16.72	0.373	9021
Mother complete secondary	12.94	0.336	9021
Mother some tertiary	6.61	0.249	9021
Non-Muslim	7.39	0.261	9021

Household circumstances

We have also included indicators of household living conditions as further circumstances potential associated with student performance. First, we construct a simple index of durable goods which adds up the ownership of radio, video recorder, music player, telephone, motorbike, computer, and refrigerator. Then we generate the following dummies: (1) household has one of the mentioned durable goods; (2) household has two durables; (3) household has three durables; (4) household has four durables; (5) household has five or more durables. The omitted category is not having any of those durable goods. We proceed in a similar way to count the number of news outlets read in the house. We add up purchases of newspapers and magazines and define the dummies: (1) household has one of the two items; (2) household has both newspapers and magazines. The omitted category is having none of them.

Additionally we include a dummy taking the value of 1 if the household does not own any farming livestock, and a dummy taking the value of 1 if there is arsenic in the water supply. The descriptive statistics for household circumstances are in table 3.7.

	Mean (%)	Standard deviation	Sample size
One durable	20.39	0.403	9021
Two durables	21.03	0.408	9021
Three durables	18.84	0.391	9021
Four durables	12.64	0.332	9021
Five or more durables	10.25	0.303	9021
One news item	26.82	0.443	9021
Two news items	10.17	0.302	9021
No farming livestock	33.89	0.473	9021
Arsenic in water	50.21	0.500	8493

Table 3.7 Household characteristics

School characteristics

Finally we control for any potential source of variation in scores associated with differences between schools by adding school fixed effects, as we have information on the particular school attended by each student, in addition to detailed information on these schools' characteristics. However, we are not interested in unpacking the specific features that are more associated with test performance, i.e. we are not estimating the parameters of an "education production function", we just control for between-school variation in the aforementioned manner. We use a vector of five dummy variables to measure school-specific traits including whether the school is a madrasah Aliya, whether the secondary school admits children from any primary school, whether the school has electricity, has a library, and has a computer. In our analysis we do not include the minority of students who attend unregistered madrasas.

Table 3.8 Type of school attended

	Mean (%)	Standard deviation	Sample size
Registered madrasah	26.43	0.441	9021
Admission from any school	66.12	0.473	9021
Access to electricity	84.90	0.358	9021
Access to a library	56.62	0.495	9021
Access to a computer	52.07	0.500	9021

4. Methodology

a. Estimation strategies

We follow the proposal by Jusot et al. (2013), but with minor adjustments due to special features of our dataset. Since we want to gauge the empirical relevance of the different views as to how the correlation between efforts and circumstances should affect the magnitude of legitimate and illegitimate sources of inequality in education in rural Bangladesh, we model the variation of indicators of educational achievement in mathematics (A^m) and in English language (A^l) , as functions of a vector of students' demographic variables (D), a vector of their circumstances (C), a vector of school characteristics (F), a vector of efforts in mathematics (E^m) only included in the equation for mathematics achievement, and a vector of efforts in English (E^l) only included in the equation for English language achievement. For each equation there is also an error term $(u^m$ and u^l):

$$A^m = f(D, C, F, E^m, u^m) \tag{1}$$

$$A^{l} = f(D, C, F, E^{l}, u^{l})$$

$$\tag{2}$$

Since the dependent variables are deemed continuous, we can estimate (1) and (2) with a linear model. As mentioned in the data section, D includes age, and sex, i.e. circumstances unrelated to family characteristics. C includes parental and household circumstances: the dummies for father's and mother's education, the dummies for number of durable goods, dummies for number news outlets, the dummy for arsenic in the water, and the dummy for lack of farming livestock, along with the religious affiliation. F includes the set of school-specific dummies that control for school fixed effects. Finally E^m and E^l contain the effort dummies described in the data section for the mathematics and English test scores, respectively.

Following Jusot et al. (2013) we deem the contribution of parental and household circumstances to total inequality illegitimate as inequality of opportunity, as opposed to legitimate inequality due to effort. However in the light of the previous discussion involving the views of Roemer and Barry, we need separate estimation strategies reflecting how each view treats the correlation between circumstances and efforts.

The Barry case

In the case of Barry's view, performance differences due to students' efforts would need to be fully respected. Therefore, given the nature of our outcome variables we can estimate the following two equations:

$$A_{i}^{m} = \lambda^{m,B} + \alpha^{m,B}C_{i} + \beta^{m,B}E_{i}^{m} + \gamma^{m,B}D_{i} + {}^{m,B}F_{s} + u_{i}^{m}$$
(3)

$$A_i^l = \lambda^{l,B} + \alpha^{l,B}C_i + \beta^{l,B}E_i^l + \gamma^{l,B}D_i + \delta^{l,B}F_s + u_i^l$$

$$\tag{4}$$

where the *i* subscripts represent students' individual values for the variables in the respective vectors, and the Greek letters superscripted by *B* for Barry are coefficients. In the linear setting of (3) and (4) we would declare equality of opportunity in maths scores according to Barry's view if $\alpha^{m,B} = 0$. Likewise, for English scores we would require: $\alpha^{l,B} = 0$.

The school fixed-effects enable us to capture as much as possible of the part of the explained variation in the outcome attributable to *between-school effects* (in terms of association rather than causation). Yet for this very reason these fixed effects are a sort of "black box" since between-school fixed effects may be embodying effects related to: school quality differences, between-school circumstances (e.g. affecting school choice), and between-school effort differentials, and between-school demographics. We will discuss a proposal to unpack this below. In the meantime it is worth checking how equations (3) and (4) look when we try to estimate the parameters of the explanatory variables other than the school dummies:

$$A_{is}^{m} - A_{s}^{m} = \alpha^{m,B}(C_{i} - C_{s}) + \beta^{m,B}(E_{i}^{m} - E_{s}^{m}) + \gamma^{m,B}(D_{i} - D_{s}) + u_{i}^{m} - u_{s}^{m}$$
(5)

$$A_{is}^{l} - A_{s}^{l} = \alpha^{l,B}(C_{i} - C_{s}) + \beta^{l,B}(E_{i}^{l} - E_{s}^{l}) + \gamma^{l,B}(D_{i} - D_{s}) + u_{i}^{l} - u_{s}^{l}$$
(6)

where A_s^m , A_s^l , C_s , E_s^m , E_s^l , D_s represent school-level averages for the respective sets of variables. Essentially, equations (5) and (6) show us that with this current model, we are mainly capturing the associated effects of *within-school variations* in efforts, circumstances, and demographic characteristics on within-school variations in test scores.

Now we can model, in parallel, between-school variations in the two outcomes, as function of between-school differences in circumstances, efforts, and school quality (denoted by the vector of school-specific traits Q_s):

$$A_{s}^{m} = \pi^{k,m} + \zeta^{m,B}C_{s} + \eta^{m,B}E_{s}^{m} + v^{m,B}D_{s} + \omega^{m,B}Q_{s} + \epsilon_{s}^{m}$$
⁽⁷⁾

$$A_{s}^{l} = \pi^{k,l} + \zeta^{l,B}C_{s} + \eta^{l,B}E_{s}^{l} + \upsilon^{l,B}D_{s} + \omega^{l,B}Q_{s} + \epsilon_{s}^{l}$$
(8)

Given the nature of our educational outcome variables we can use their predicted values from the linear models above as linearly decomposable measures of educational attainment:

$$(A_{is}^{\widetilde{k}} - A_{s}^{k}) = \alpha^{\widetilde{k},B}(C_{i} - C_{s}) + \beta^{\widetilde{k},B}(E_{i}^{m} - E_{s}^{m}) + \gamma^{\widetilde{k},B}(D_{i} - D_{s})$$

$$(9)$$

$$\widetilde{A_s^k} = \pi^{\widetilde{k},B} + \zeta^{\widetilde{k},B}C_s + \eta^{\widetilde{k},B}E_s^l + v^{\widetilde{k},B}D_s + \omega^{\widetilde{l},B}Q_s \tag{10}$$

Where $k = m, l, (A_{ls}^{k} - A_{s}^{k})$ are the predicted deviations of maths and English scores for each individual "i" from their respective school means, under the Barry (B) view, while \widetilde{A}_s^k are the predicted average scores in school "s". The accented coefficients are the estimates from each respective model. Then, in order to decompose the inequality in these indicators into legitimate and illegitimate components, we follow Jusot et al. (2013) and measure absolute inequality with the variance and/or relative inequality with the squared coefficient of variation, since these are the only inequality measures which are linearly decomposable by sources and fulfil a set of desirable decomposition properties (Shorrocks, 1982). Since the square coefficient of variation is just the variance divided by the squared mean, then the decomposition for both is the same. We use the variance, which belongs in a class of additively decomposable absolute inequality index (Bosmans and Cowell, 2010), in the sense that it can be decomposed into within-group and between-group components in the same way that Shorrocks (1980) showed for a class of relative inequality indices. Moreover, the group decomposition of the variance satisfies a property of path-independence (Foster and Shneyerov, 2000) whereby the within-group component is merely a population-share-weighted average of the within-group variances, whereas the betweengroup variance provides a measure of dispersion for a smoothed distribution in which each individual observation has been replaced by its respective group average.

Let $\widetilde{P_{ts}^k} \equiv (A_{ts}^k - A_s^k) + \widetilde{A_s^k}$ be the predicted individual score, assembled indirectly from the two models. Then the variance of $\widetilde{P_{ts}^k}$ can be decomposed in the following way:

$$\sigma^{2}(\widetilde{P_{ls}^{k}}) \equiv \frac{1}{N} \sum_{i=1}^{N} \left(\widetilde{P_{ls}^{k}} - \widetilde{P^{k}} \right)^{2} = \sigma_{BS}^{2}(\widetilde{P_{ls}^{k}}) + \sigma_{WS}^{2}(\widetilde{P_{ls}^{k}})$$
(11)

where N is the number of children, $\widetilde{P^k}$ is the average predicted score across all children, $\sigma_{BS}^2(\widetilde{P_{ls}^k})$ is the between-school component of the variance, and $\sigma_{WS}^2(\widetilde{P_{ls}^k})$ is the within-school component. Each in turn is defined by:

$$\sigma_{BS}^2(\widetilde{P_{lS}^k}) \equiv \sum_{s=1}^{S} n_s \left(\widetilde{P_{S}^k} - \widetilde{P^k}\right)^2 \tag{12}$$

$$\sigma_{WS}^2(\widetilde{P_{ls}^k}) \equiv \sum_{s=1}^{S} n_s \frac{1}{n_s N} \sum_{i=1}^{n_s N} \left(\widetilde{P_{ls}^k} - \widetilde{P_s^k} \right)^2 \tag{13}$$

where n_s is the proportion of children in school "s", and \widetilde{P}_s^k is the average predicted score in school "s" (note that it may differ from \widetilde{A}_s^k). Finally, we define the contributions by each component to the variance:

$$\theta_{BS} \equiv \frac{\sigma_{BS}^2(P_{lS}^k)}{\sigma^2(\bar{P}_{lS}^k)} \tag{14}$$

$$\theta_{WS} \equiv \frac{\sigma_{WS}^2(\bar{P}_{LS}^{\bar{k}})}{\sigma^2(\bar{P}_{LS}^{\bar{k}})} \tag{15}$$

We will use θ_{BS} and θ_{WS} below in order to compute the contributions of effort and circumstances to the total predicted variation in the scores. First we show how we derive the contributions of circumstances and efforts to each of the variance components.

Let $C_{l}^{\widetilde{k},B} \equiv \alpha^{\widetilde{k},B}(C_{l} - C_{s})$ and $C_{s}^{\widetilde{k},B} \equiv \zeta^{\widetilde{k},B}C_{s}$ be the parts of the predicted score attributable to circumstances in the within-school and the between-school models, respectively, both in the Barry case (and similar definitions for the other elements). Then the decomposition of the variance of the predicted scores within any (one) given school is given by:

$$\sigma_s^2((A_{\iota s}^{\widetilde{k}} - A_s^k)) = cov\left((A_{\iota s}^{\widetilde{k}} - A_s^k), \widetilde{C_\iota^{\widetilde{k}, B}}\right) + cov\left((A_{\iota s}^{\widetilde{k}} - A_s^k), \widetilde{E_\iota^{\widetilde{k}, B}}\right) + cov\left((A_{\iota s}^{\widetilde{k}} - A_s^k), \widetilde{D_\iota^{\widetilde{k}, B}}\right)$$
(16)

Where k = m, l. Meanwhile, the decomposition of the variance of the predicted scores between schools is given by:

$$\sigma_{BS}^{2}(\widetilde{A}_{s}^{k}) = cov\left(\widetilde{A}_{s}^{k}, \widetilde{C}_{s}^{k,B}\right) + cov\left(\widetilde{A}_{s}^{k}, \widetilde{E}_{s}^{k,B}\right) + cov\left(\widetilde{A}_{s}^{k}, \widetilde{D}_{s}^{k,B}\right) + cov\left(\widetilde{A}_{s}^{k}, \widetilde{Q}_{s}^{k,B}\right)$$
(17)

The contribution of circumstances to overall explained within-school variability in one school in the Barry view is given by:

$$cov\left((A_{ls}^{\widetilde{k}}-A_{s}^{k}), \widetilde{C_{l}^{k,B}}\right) = \sigma^{2}\left(\widetilde{C_{l}^{k,B}}\right) + \rho_{CE}\sigma\left(\widetilde{C_{l}^{k,B}}\right)\sigma\left(\widetilde{E_{l}^{k,B}}\right) + \rho_{CD}\sigma\left(\widetilde{C_{l}^{k,B}}\right)\sigma\left(\widetilde{D_{l}^{k,B}}\right)$$
(18)

Where ρ_{CE} is the correlation coefficient between circumstance and effort parts of the predicted score (and same definition for ρ_{CD} , etc.). Meanwhile, the contribution of efforts is given by:

$$cov\left((\widetilde{A_{\iota S}^{k} - A_{S}^{k}}), \widetilde{E_{\iota}^{k,B}}\right) = \sigma^{2}\left(\widetilde{E_{\iota}^{k,B}}\right) + \rho_{CE}\sigma\left(\widetilde{C_{\iota}^{k,B}}\right)\sigma\left(\widetilde{E_{\iota}^{k,B}}\right) + \rho_{ED}\sigma\left(\widetilde{E_{\iota}^{k,B}}\right)\sigma\left(\widetilde{D^{k,B}}\right)$$
(19)

Likewise, for the between-school model we have the respective contributions of circumstances and efforts (where $\rho_{CE(S)}$ is the correlation coefficient between circumstance and effort parts of the predicted score in the between-school model, and same definition for the other coefficients) :

$$cov(\widetilde{A}_{s}^{k},\widetilde{C}_{s}^{\widetilde{k},B}) = \sigma^{2}(\widetilde{C}_{s}^{\widetilde{k},B}) + \rho_{CE(S)}\sigma(\widetilde{C}_{s}^{\widetilde{k},B})\sigma(\widetilde{E}_{s}^{\widetilde{k},B}) + \rho_{CD(S)}\sigma(\widetilde{C}_{s}^{\widetilde{k},B})\sigma(\widetilde{D}_{s}^{\widetilde{k},B}) + \rho_{CQ(S)}\sigma(\widetilde{C}_{s}^{\widetilde{k},B})\sigma(\widetilde{Q}_{s}^{\widetilde{k},B})$$
(20)

$$cov\left(\widetilde{A_{s}^{k}},\widetilde{E_{s}^{k,B}}\right) = \sigma^{2}\left(\widetilde{E_{s}^{k,B}}\right) + \rho_{CE}\sigma\left(\widetilde{C_{s}^{k,B}}\right)\sigma\left(\widetilde{E_{s}^{k,B}}\right) + \rho_{ED}\sigma\left(\widetilde{E_{s}^{k,B}}\right)\sigma\left(\widetilde{D^{k,B}}\right) + \rho_{EQ(S)}\sigma\left(\widetilde{E_{s}^{k,B}}\right)\sigma\left(\widetilde{Q_{s}^{k,B}}\right)$$
(21)

Then, in order to compute the total within-school variance we add $\sigma_s^2((A_{is}^k - A_s^k))$ across schools weighting by relative school size n_s :

$$\sigma_{within}^2 = \sum_{s=1}^{S} n_s \sigma_s^2 \left((A_{is}^{\widetilde{k}} - A_s^k) \right) \tag{22}$$

Accordingly the respective total contributions of effort and circumstances to the total withinschool variation are:

$$cov_{within}\left((A_{ls}^{\widetilde{k}}-A_{s}^{k}), \widetilde{C_{l}^{\widetilde{k},B}}\right) = \sum_{s=1}^{S} n_{s} cov\left((A_{ls}^{\widetilde{k}}-A_{s}^{k}), \widetilde{C_{l}^{\widetilde{k},B}}\right)$$
(23)

$$cov_{within}\left((\widetilde{A_{ls}^{k}-A_{s}^{k}}),\widetilde{E_{l}^{k}}\right) = \sum_{s=1}^{S} n_{s}cov\left((\widetilde{A_{ls}^{k}-A_{s}^{k}}),\widetilde{E_{l}^{k}}\right)$$
(24)

Using the above equations, we define the contributions of circumstances and efforts, respectively, to total within-group variance and to between-group variance as follows:

$$\tau_{WS}^{C} \equiv \frac{cov_{within}\left(\left(A_{ls}^{\widetilde{k}-A_{S}^{k}}\right), \widetilde{c_{l}^{k,B}}\right)}{\sigma_{within}^{2}}$$
(25)

$$\tau_{WS}^{E} \equiv \frac{cov_{within}\left(\left(A_{ls}^{\widetilde{k}-A_{s}^{k}}\right), \widetilde{E_{l}^{k,B}}\right)}{\sigma_{within}^{2}}$$
(26)

$$\tau_{BS}^{C} \equiv \frac{cov(\widetilde{A_{s}^{k}}, \widetilde{c_{s}^{k,B}})}{\sigma_{BS}^{2}(\widetilde{A_{s}^{k}})}$$
(27)

$$\tau_{BS}^{C} \equiv \frac{cov(\widetilde{A_{s}^{k}}, \widetilde{E_{s}^{k,B}})}{\sigma_{BS}^{2}(\widetilde{A_{s}^{k}})}$$
(28)

Likewise we can also define the contributions of demographic characteristics and school quality variables. Finally, we define and compute the total contributions of circumstances and efforts, respectively, the following way:

$$\theta_C \equiv \theta_{BS} \tau_{BS}^C + \theta_{WS} \tau_{WS}^C \tag{29}$$

$$\Theta_E \equiv \theta_{BS} \tau_{BS}^E + \theta_{WS} \tau_{WS}^E \tag{30}$$

The Roemer case

In Roemer's view we would only need to respect differences due to effort, which, in turn, cannot be attributed to circumstances. Hence, as a first step, Jusot et al. (2013) proposed using a set of auxiliary equations in which the effort variables are modelled as a function of the circumstances:

$$E_{is}^{m} = \xi^{m} + \mu^{m,R}C_{i} + \nu^{m,R}D + \phi^{m,R}F_{s} + e_{i}^{m}$$
(31)

$$E_{is}^{l} = \xi^{l} + \mu^{l,R}C_{i} + \nu^{l,R}D + \phi^{l,R}F_{s} + e_{i}^{l}$$
(32)

where e_i^m and e_i^l are vectors of residual terms. Since our efforts are binary variables then we estimate (31) and (32) with Probit models. Then as a second step, they prescribe replacing E_i^m and E_i^l in (31) and (32) with the estimated vector of residuals from (5) and (6), i.e. \tilde{e}_i^m and \tilde{e}_i^l . These are actually generalised residuals stemming from a non-linear model.⁷

The procedure then yields:

$$A_i^m = \lambda^{m,R} + \alpha^{m,R}C_i + \beta^{m,R}\widetilde{e_i^m} + \gamma^{m,R}D + \delta^{m,R}F + u_i^m$$
(33)

$$A_i^l = \lambda^{l,R} + \alpha^{l,R}C_i + \beta^{l,R}\widetilde{e_i^l} + \gamma^{l,R}D + \delta^{l,R}F + u_i^l$$
(34)

Finally we declare equality of opportunity in Roemer's view if $\alpha^{m,R} = 0$ and $\alpha^{l,R} = 0$, for the maths and English scores, respectively.

In our framework we need to do these replacements both in the within-school and the betweenschool models. The rest of the procedure proceeds the same way as prescribed in the previous subsection.

5. Results

a. Mathematics and English scores equations in the two viewpoints

Table A1 in the appendix shows the results for equations (5) and (6), i.e. for the models reflecting Barry's view within-schools. Table A2 shows the results for corresponding to Roemer's view. Finally tables A3 and A4 show the respective auxiliary equations, i.e. (31) and (32), in which we model students' efforts as functions of the other circumstances (including parental, household, school effects, and own demographic characteristics), as a prerequisite procedure for the implementation of the model reflecting Roemer's position.

In terms of mathematics under Barry's view, table A1 shows that all effort variables are positively and significantly associated with the score except incomplete homework and absenteeism. Parental education dummies are positively related to the score in that any level is associated with better scores vis-a-vis the omitted category of less than complete primary education with the exception of mother with complete primary education. However only in the case of mother's education extra attainment levels systematically come along with higher scores. While all dummies are statistically significant among fathers only the dummy of higher education is

⁷ See Appendix A in Jusot et al. (2013) for the technical details.

significant for mother's education. Meanwhile among the asset variables, none of them are significant and only the dummy for five or more durables is positively related to the score. Having more than two news outlets is positively and significantly associated with the score. The presence of arsenic and being from a religious minority are negatively but not significantly associated with the score. Regarding individual demographic variables, being a female student significantly decreases by more than one point the mathematics score, and higher age is increasing negatively associated with mathematics score with a significant negative association between the oldest age group with the score.

Regarding English language score, all effort dummies are statistically significant and positively related to the score; the only exception is the residual effort related to disinterest in schoolwork being non significant at the 10% level. Both father and mother's education levels are positively associated with higher scores and the relationships are monotonic; while the associations are significant for all the mother's dummies, they are only significant above complete secondary school for the father's education. Most other household characteristics are not statistically significant with the exception of not owning livestock (positive relationship), arsenic in the water (negative relationship) and being from a religious minority (negative relationship). Likewise most individual demographic variables including gender are negatively and statistically significant with the exception of one age dummy.

Table A2 shows the results for the model of both scores under Roemer's view. The results are remarkably similar to the results in the Barry case. The same residual effort variables are positively and significantly associated with the score (except noise in the classroom) and as expected the estimated coefficients are deflated compared to the Barry case. On the other hand, the estimated coefficients of the variables labelled as circumstances including parental education dummies, household characteristics are larger than in the Barry case. As observed in the Barry case parental education dummies all have positive coefficients (except mother's having completed primary) meaning that the respective education levels are associated with higher score vis-a-vis the omitted category of less than complete primary education and the significance of the relationship between parental education levels and the offspring's mathematics score is stronger in Romer's context. As for other household variables, more than five durables and more than two news items is significantly and positively associated with the score. While the lack of livestock was not significantly associated with the mathematics score in the Barry context, it is positively and significantly associated at 10% with the score in the Roemer case. Like Barry's context, being a female and in the oldest age group is negatively and significantly associated with the score.

In terms of English score, again the results are remarkably the same in the Roemer case as they were in the Barry case in terms of significance and sense of the associations. Again, we note deflated estimated coefficients for all residual effort variables while variables related to circumstances (parental education, number of durables, news items and other household characteristics) exhibit larger estimated coefficients than in the Barry case.

Table A3 shows the auxiliary equations for each of the seven effort dummies at mathematics as functions of all the other "circumstance" variables. Marginal effects are presented. There are few consistent patterns in terms of marginal effects of the same variable across different efforts as dependent variables. However, interestingly, all the father's education dummies bear positive marginal effects (albeit not always statistically significant at 10%) across equations. Table A4 shows the auxiliary equations for each of the seven effort dummies at English as functions of all the other "circumstances". As in the case of the mathematics auxiliary equations, few consistent patterns of marginal effects can be found across equations; however, all the father's education

dummies, again, have positive marginal effects. Given the nature of the omitted educational category, these effects mean that having complete primary education or more is associated with better efforts.

Table A5 shows the between-schools regression results for each score in Barry's view. It is noticeable that those estimations are undertaken for the purpose of estimating the between school variance and are undertaken on a small sample of observations (262 schools for the Mathematics score and 284 schools for the English score) and therefore few significant associations are observed. For the mathematics score, the only significant effort between schools is students' making noise rarely being positively associated with the score. The mother's education dummies above complete primary education are positively and significantly associated with the score and the association is increasing with education extra attainment levels. Between schools variation in mathematics score are negatively and significantly associated with some household characteristics including availability of two durables and availability of four durables and above, and lack of livestock ownership. Within the school-specific characteristics, only availability of a computer in the school is positively and significantly associated with the mathematics score. As for the English score, none of the efforts are significantly associated with the score with the exception of absenteeism, which is positively and significantly associated with the English score between schools. The highest dummy of mother's education is positively and significantly associated with higher English scores, as is the father's lowest dummy of education (incomplete primary). Among household characteristics, only the availability of two durables is and the availability of two news outlets or more are found significantly associated with the English score (negative association for durables and positive association for the news outlet). Finally a number of school traits are significantly associated with the variance in English score: being a Madrasa and having electricity are negatively associated with the score, while the availability of a computer is positively associated with the score.

Table A6 shows the results for the between-schools regressions model of mathematics and English scores under Roemer's view. The results are remarkably similar to the results in the Barry case for both scores. The only exception is a change in the significance of the association between the mathematics score and students' making noise rarely now just above the 10% significance level. As underlined in the within-school estimations, Roemer's view leads to deflated estimated coefficients for the residuals variables and inflated estimated coefficients for the circumstances-related variables.

b. Relative contributions to educational inequality

Table 5.1 shows the results of the decomposition exercise laid out in the previous section. Each row shows the relative contribution of each vector of variables to the within-schools and between schools predicted variances of scores for mathematics and English, under the Barry and the Roemer scenarios, in their respective columns. The rows labelled "variance" show the variance of the predicted scores for each combination of subject and within or between schools contexts.

Table 5.1 Decom	position of	f educational	inequalities	by source
-----------------	-------------	---------------	--------------	-----------

	Mathem	atics	Engli	sh
	Barry	Roemer	Barry	Roemer
TOTAL VARIANCE				
Efforts (%)	20.73	18.28	30.58	27.13
Circumstances (%)	47.02	50.89	40.27	43.49
Demographics (%)	19.91	20.89	5.27	4.94
School traits (%)	12.34	9.94	23.88	24.43
WITHIN SCHOOLS variance	0.507		0.600	5
Total share of within schools inequality	20.37	21.01	21.55	20.00
Efforts (%)	6.95	5.96	14.94	12.51
Circumstances (%)	2.42	3.09	5.07	5.80
Demographics (%)	10.99	11.96	1.54	1.69
BETWEEN SCHOOLS variance	1.983	3	2.20	5
Total share of between schools inequality	79.65	78.99	78.45	80.00
Efforts (%)	13.78	12.33	15.64	14.62
Circumstances (%)	44.59	47.79	35.20	37.69
Demographics (%)	8.91	8.93	3.74	3.25
School traits (%)	12.34	9.94	23.88	24.43

Firstly, we note that the degree of legitimate inequality (effort-related) varies by subject, for each particular view, and is significantly higher for English test scores. For Barry's view, the degree of legitimate inequality rises from 20.73% in the case of mathematics to 30.58% in the case of English. As expected, in Roemer's view the magnitude of legitimate inequality is smaller, but still contributes for 18.28% in the variance in mathematics score and 27.13% in English score. Interestingly, while the decrease in the share of legitimate inequality when we adopt Roemer's view is proportionally similar for both scores, the absolute discrepancy between the two views is larger in the case of English.

The other side of the coin is the increase in the contribution of circumstances when moving from Barry's to Roemer's view. Noticeably the contribution of circumstances is the largest in the total variance of each score regardless of the viewpoint, it ranges from 40.27% (in Barry's view applied to English) to more than half of the explained variance of the mathematics score in Roemer's view (50.89%).

The relative importance of demographics and reciprocally the relative importance of school traits strikingly depend on the subject. In the case of mathematics, demographic characteristics are as important as effort and represents one fifth of the inequality, whereas they make the smallest contribution to inequalities in English score with a contribution of 5% on average in each viewpoint. On the other hand, school traits contribute for 23.88% in the variance in English score and 12.34% in mathematics score inequalities in Barry's view. While in the case of English, the contribution of school traits follow the pattern of circumstances and increase its contribution to the total explained variance when we adopt Roemer's view; its contributions in the case of mathematics score reduces to 9.94%.

The breakdown of the contributions of each source into within-schools and between schools context underline that circumstances matter more for between-school variations and efforts matter more for within-school variations for both subjects. The importance of the viewpoint adopted appears to matter more in the within-school models than in the between-schools, especially for mathematics for which the contribution of circumstances increases by 22% (respectively 13% for the English score) and the contribution of effort decreases by 16% (respectively 19% for the English score).

6. Concluding remarks

Jusot et al. (2013) were the first to propose a straightforward method to quantify the contribution of legitimate and illegitimate inequalities (i.e. inequality of opportunity), accommodating different views on how the correlation between efforts and circumstances should be considered when defining the boundaries between the two sources of inequality. Since their empirical application was on health outcomes in France, the authors were recommending, in one of their concluding remarks, replicating their method using different health variables. They were also emphasizing the importance of counting on richer datasets with more information on outcomes, circumstances, and mainly effort indicators.

In this paper we have applied their method to the measurement of legitimate and illegitimate inequality in educational outcomes of secondary-school children in rural Bangladesh (admittedly a significantly different setting). We found that the relative contribution of each form of inequality does not change significantly when we move from Barry's to Roemer's view, in the sense that, under both scenarios, the degree of legitimate inequality is relatively small, and significantly lower than the illegitimate counterpart, which account for about half of the inequalities in education. However, as expected the contribution of circumstances is higher, and that of effort lower, under Roemer's view since it requires purging out any variation in effort attributable to circumstances.

On the other hand, we found corroboration that the extent of inequality of opportunity is contextual, not only across different dimensions of wellbeing (e.g. health versus education) but also across wellbeing indicators *within* a particular dimension.

Our results underlined substantial gender and age differences by subject, it would be interesting to investigate further this result and look further into studies on the relationship between gender and age with cognitive abilities particularly in mathematics.

Finally we must emphasize that a recurring empirical challenge in this and related decomposition methods is the availability of indicators for all the required types of variables, i.e. outcomes, efforts, and circumstances. The relative abundance of effort indicators vis-à-vis circumstances may affect their respective contributions.

7. References

Arneson, R. J. 1990. "Liberalism, Distributive Subjectivism, and Equal Opportunity for Welfare." *Philosophy and Public Affairs* 19, 158-194.

Asadullah, N., Chaudhury, N., and A. Dar (2007) "Student achievement conditioned upon school selection: Religious and secular secondary school quality in Bangladesh", *Economics of Education Review*, 26: 648-59.

Asadullah, N., Chaudhury, N. and S. Josh, (2009) "Secondary school Madrasas in Bangladesh: Incidence, Quality, and Implications for Reform", The World Bank, Human Development Sector, South Asia Region.

Asadullah, N. and N. Chaudhury (2009) "Holy alliances: public subsidies, Islamic high schools, and female schooling in Bangladesh", *Education Economics*, 17(3): 377-94.

Asadullah, N., Chakrabarti, R., and N. Chaudhury, (2013) "What determines religious school choice? Theory and evidence from rural Bangladesh", Bulletin of Economic Research, DOI: 10.1111/j.1467-8586.2012.00476.x

Asadullah, N. (2014) "The effect of Islamic secondary school attendance on academic achievement", IZA DP 8233.

Barry, B. (2005) Why social justice matters. Cambridge: Polity Press.

Bjorklund, A., and K. Salvanes (2011), "Education and family background: mechanisms and policies", chapter 3 in Hanushek, E. et al. (editors), *Handbook of the Economics of Education*, Volume 3: 201-47, Elsevier.

Bosmans, K. and F. Cowell (2010) "The class of absolute decomposable inequality measures", *Economic Letters*, 109: 154-156.

Checci, D. and V. Peragine (2010) "Inequality of opportunity in Italy", *Journal of Economic Inequality*, 8: 429-50.

De Fraja, G., Oliveira, T., and L. Zanchi (2010) "Must try harder: Evaluating the role of effort in educational attainment", *The Review of Economics and Statistics*, 92(3): 577-97.

Eren, O., and D. Henderson (2011) "Are we wasting our children's time by giving them more homework?", *Economics of Education Review*, 30(5): 950-61.

Ermisch, J, and M. Francesconi (2001) "Family matters: Impacts of family background on educational attainments", *Economica*, 68: 137-56.

Feinstein, L. (2003) "Inequality in the Early Cognitive Development of British Children in the 1970 Cohort", *Economica*, 70(277): 73-97.

Ferreira, F. and J. Gignoux (2011) "The measurement of inequality of opportunity: Theory and application to Latin America", *Review of Income and Wealth*, 57(4): 622-57.

Fleurbaey, M. and V. Peragine (2013) "Ex ante versus ex post equality of opportunity", *Economica*, 80(317): 118-30.

Fleurbaey, M. and E. Schokkaert (2009) "Unfair inequalities in health and health care", *Journal of Health Economics*, 28: 73-90.

Foster, J. and A. Shneyerov (2000) "Path independent inequality measures", *Journal of Economic Theory*, 91: 199-222.

Garcia-Gomez, P., Schokkaert, E., Van Ourti, T., and T. Bago D'Uva (2014) "Inequality in the face of death", *Health Economics*, DOI: 10.1002/hec.3092

Haveman, R., and B. Wolfe (1995) "The determinants of children's attainments: A review of methods and findings", *Journal of Economic Literature*, 33(4): 1829-78.

Jusot, F., Tubeuf S. and A. Trannoy (2013) "Circumstances and efforts: how important is their correlation for the measurement of inequality of opportunity in health?", *Health Economics*, 22(12): 1470-95.

Kuen, Z., and P. Landeras (2014) "The effect of family background on student effort", *The B.E. Journal of Economic Analysis & Policy*, 14(4): 1337-1403.

Lefranc, A., Pistolesi, N., and A. Trannoy (2008) "Inequality of opportunities versus inequality of outcomes: Are western societies all alike?" *Review of Income and Wealth*, 54(4): 513-46.

Lefranc, A., Pistolesi, N., and A. Trannoy (2009) "Equality of opportunity and luck: Definitions and testable conditions, with an application to income in France", *Journal of Public Economics*, 93(11-12): 1189-207.

Lu, I., Chanel.O., Luchini S. and A.Trannoy 2013; "Responsibility cut in Education and Income acquisition: An empirical investigation" Amse WP n°2013-47.

Metcalfe, R., Burgess S., and S. Proud (2011) "Student effort and educational attainment: Using the England football team to identify the education production function", CMPO Working Paper 11/276.

Ooghe, E., Schokkaert, E., and D. van de Gaer (2007) "Equality of opportunity versus equality of opportunity sets", *Social Choice and Welfare*, 28: 209-30.

Roemer J. Equality of opportunity. Harvard University Press; Cambridge; 1998.

Roemer, J., and A. Trannoy (2015) "Equality of opportunity", chapter 4 in Atkinson, A. et al. (editors), *Handbook of Income Distribution*, Volume 2: 217-300, Elsevier.

Shorrocks, A. (1980) "The class of additive decomposable inequality measures", *Econometrica*, 48(3): 613-625.

Shorrocks, A. (1982) "Inequality decomposition by factor components", *Econometrica*, 50(1): 193-211.

Stinebrickner, R. and T. Stinebrickner (2008) "The casual effect of studying on academic performance", The B.E. Journal of Economic Analysis & Policy, 8(1): 1935-1982.

US Department of Education (1988) National Education Longitudinal Study of 1988. Teacher questionnaire NELS 88:Base year.

Van de Gaer, D. (1993) Equality of opportunity and investment in human capital, KULeuven, Leuven.

8. Appendix

Explanatory variables	Dependent va	riable:	Dependent variable:	
	Mathematics score		English s	score
	Coefficient	P-value	Coefficient	P-value
Perform below ability (never/rarely)	0.348	0	0.801	0
Homework (never/rarely)	0.204	0.014	0.364	0
Absent (never/rarely)	0.259	0.002	0.486	0
Lazy (never/rarely)	0.210	0.017	0.470	0
Inattentive (never/rarely)	0.337	0	0.265	0.005
Disinterested (never/rarely)	0.280	0.002	0.102	0.417
Noisy (never/rarely)	0.249	0.039	n/a	n/a
Mother complete primary	-0.012	0.89	0.193	0.049
Mother some secondary	0.125	0.193	0.417	0
Mother complete secondary	0.145	0.175	0.425	0
Mother some tertiary	0.294	0.037	0.431	0.004
Father complete primary	0.181	0.087	-0.007	0.953
Father some secondary	0.162	0.09	0.054	0.598
Father complete secondary	0.216	0.026	0.269	0.011
Father some tertiary	0.323	0.002	0.453	0
One durable good	-0.022	0.831	0.060	0.592
Two durable goods	0.107	0.307	0.070	0.535
Three durable goods	0.027	0.806	0.150	0.198
Four durable goods	-0.018	0.883	0.096	0.471
Five or more durable goods	0.307	0.022	0.035	0.812
One news item	0.075	0.336	0.048	0.573
Two news items	0.257	0.028	-0.029	0.823
No livestock	0.086	0.221	0.279	0
Arsenic	-0.024	0.784	-0.250	0.008
Non Muslim	-0.161	0.192	-0.288	0.033
13 years old	-0.004	0.959	0.004	0.967
14 years old	-0.037	0.698	-0.409	0
15 years old or older	-0.608	0.006	-0.826	0.001
Female	-1.212	0	-0.110	0.145
Number of observations	6,369		7,180)
Adjusted R-squared	0.0782		0.071	3

*Models include constant term.

Explanatory variables	Dependent variable:		Dependent va	ariable:
	Mathematics score		English score	
	Coefficient	P-value	Coefficient	P-value
Residual - Perform below ability (never/rarely)	0.217	0	0.458	0
Residual - Homework (never/rarely)	0.116	0.019	0.212	0
Residual - Absent (never/rarely)	0.142	0.004	0.266	0
Residual - Lazy (never/rarely)	0.105	0.039	0.268	0
Residual - Inattentive (never/rarely)	0.180	0	0.140	0.012
Disinterested (never/rarely)	0.139	0.007	-0.030	0.646
Residual - Noisy (never/rarely)	0.055	0.36	n/a	n/a
Mother complete primary	-0.001	0.987	0.240	0.015
Mother some secondary	0.157	0.103	0.454	0
Mother complete secondary	0.126	0.239	0.436	0
Mother some tertiary	0.347	0.014	0.478	0.002
Father complete primary	0.223	0.035	0.069	0.546
Father some secondary	0.230	0.016	0.124	0.227
Father complete secondary	0.267	0.006	0.326	0.002
Father some tertiary	0.437	0	0.590	0
One durable good	-0.039	0.704	0.037	0.742
Two durable goods	0.057	0.587	0.047	0.675
Three durable goods	-0.001	0.996	0.118	0.316
Four durable goods	-0.040	0.746	0.113	0.394
Five or more durable goods	0.310	0.021	0.076	0.604
One news item	0.094	0.23	0.118	0.163
Two news items	0.340	0.004	0.074	0.566
No livestock	0.116	0.101	0.313	0
Arsenic	-0.032	0.716	-0.263	0.005
Non Muslim	-0.173	0.162	-0.299	0.027
13 years old	-0.043	0.595	0.020	0.82
14 years old	-0.107	0.258	-0.435	0
15 years old or older	-0.638	0.004	-0.964	0
Female	-1.243	0	-0.152	0.045
Number of observations	6,369		7,180	
Adjusted R-squared	0.076		0.067	

71 11 40 34 11	C 11 1 1		
I able A2. Models	tor mathematics and	Hnolich fest scores	reflecting Roemer's view*
	101 maintinanco ana	Linghon test scores	Tenecung Roemer 5 view

*Models include constant term

Explanatory variables	Dependent effort variables (rarely/never)						
	Perform	Home	Absent	Lazy	Inatten-	Disin-	Noisy
	below	work			tive	terested	
	ability						
Mother complete primary	0.015	0.014	0.026*	0.005	0.008	0.028**	0.004
Mother some secondary	0.043***	0.023	0.023	0.023*	0.022	0.026**	0.018**
Mother some tertiary	0.064***	0.057*	0.040*	0.021	0.022	0.044**	0.006
Father complete primary	0.019	0.017	0.022	0.030**	0.013	0.001	0.015
Father some secondary	0.046***	0.020	0.024	0.018	0.030**	0.018	0.018**
Father some tertiary	0.077***	0.065***	0.072***	0.052***	0.039***	0.023*	0.017**
One durable good	0.002***	-0.037**	-0.019	-0.029*	-0.012	-0.001	0.004
Two durable goods	-0.019	-0.032*	-0.055***	-0.044***	-0.020	-0.035**	0.000
Three durable goods	0.012	-0.030	-0.035*	-0.027	-0.020	-0.025	0.007
Four durable goods	0.020	-0.036*	-0.013	-0.021	-0.036*	-0.014	0.004
Five or more durable goods	-0.038	-0.045**	-0.009	0.001	0.025	0.029	0.029**
One news item	0.025	0.012	-0.003	0.020*	0.013	0.018	-0.005
Two news items	0.070	0.073***	0.069***	0.049***	0.044***	0.062***	-0.006
No livestock	0.006***	0.025**	0.018	0.016	0.030***	0.044***	0.002**
Arsenic	-0.051	-0.017	0.004	-0.003	0.020	0.000	0.018
Non Muslim	-0.013***	-0.003	0.009	-0.020	-0.011	0.040**	-0.018*
13 years old	-0.024	-0.027**	-0.007	-0.029***	-0.026**	-0.009	-0.015
14 years old	-0.048*	-0.043***	-0.056***	-0.040***	-0.036***	-0.030**	-0.009
15 years old or older	0.024	-0.099***	-0.081**	-0.094***	0.008	0.039	0.012
Female	-0.005	-0.026**	0.010	-0.017*	-0.016	-0.012	-0.003
Admission any school	0.032	0.021	0.009	-0.001	0.008	0.006	0.015*
Madrasah attendance	-0.005**	-0.032**	0.006	0.004	0.028**	0.045***	0.040***
School has electricity	-0.089	-0.020	-0.075***	-0.040***	-0.061***	-0.050***	-0.024**
School has a library	0.021***	0.017	-0.012	-0.029**	-0.004	-0.028**	0.007
School has a computer	0.001	-0.040***	0.012	-0.048***	0.007	-0.033***	-0.018**
Number of observations	7727	7772	7814	7705	7723	7772	6446
Pseudo R-squared	0.0794	0.0639	0.0634	0.0720	0.059	0.0796	0.1148

Table A3: Auxiliary models for effort variables in mathematics.^a Marginal effects^b

(a) Models include constant term and school fixed effects.
(b) *=statistically significant at 10%; **=statistically significant at 5%; ***=statistically significant at 1%.

Explanatory variables	Dependent effort variables (rarely/never)					
	Perform	Home	Absent	Lazy	Inatten-	Disin-
	below	work			tive	terested
	ability					
Mother complete primary	0.033**	0.022	0.025**	0.008	0.013	-0.014
Mother some secondary	0.028*	0.022	-0.008	0.030**	0.015	-0.001
Mother some tertiary	0.046**	0.023	0.001	0.015	-0.009	0.003
Father complete primary	0.027	0.017	0.033**	0.028*	0.017	0.013
Father some secondary	0.029**	0.020	0.027**	0.001	0.021	0.010
Father some tertiary	0.049***	0.052***	0.048***	0.045***	0.055***	0.019**
One durable good	-0.014	-0.037**	0.004	-0.006	-0.023	-0.024**
Two durable goods	-0.007	-0.032*	0.001	-0.003	-0.025	-0.036***
Three durable goods	0.008	-0.017	-0.012	-0.035**	-0.020	-0.038***
Four durable goods	0.004	0.011	0.013***	0.021	-0.005	-0.039***
Five or more durable goods	0.016	-0.012	0.052***	0.029	0.003	-0.029*
One news item	0.017	0.034***	0.037***	0.044***	0.032***	0.021***
Two news items	0.034*	0.048***	0.043*	0.057***	0.061***	0.032***
No livestock	0.009	0.017	0.019	0.029*	0.014	0.002
Arsenic	-0.002	-0.005	-0.010	-0.002	-0.011	0.008
Non Muslim	-0.009	0.030	0.028	0.005	0.001	0.004
13 years old	0.012	0.015	0.017*	0.002	-0.015	0.018**
14 years old	-0.027*	-0.026*	0.024***	-0.004	-0.026*	0.010
15 years old or older	-0.056*	-0.088***	-0.085	-0.020	-0.060*	-0.014
Female	-0.015	0.016	0.004***	-0.011	-0.005	-0.033***
Admission any school	0.020	-0.005	0.063*	0.026**	0.041***	0.039***
Madrasah attendance	-0.013	-0.045***	-0.023***	0.019	0.045***	0.034***
School has electricity	-0.043***	-0.098***	-0.063	-0.065***	-0.045***	-0.030***
School has a library	0.024*	-0.001	-0.006	-0.022*	-0.040***	0.012
School has a computer	-0.033**	0.021	0.012	-0.013	-0.005	-0.020**
Number of observations	8035	8032	7975	8012	7957	7224
Pseudo R-squared	0.0568	0.0616	0.0765	0.0627	0.0658	0.1159

Table A4: Auxiliary models for effort variables in English.^a Marginal effects^b

(a) Models include constant term and school fixed effects.
(b) *=statistically significant at 10%; **=statistically significant at 5%; ***=statistically significant at 1%.

Explanatory variables	Dependent va Mathematics	riable: score	Dependent variable: English score		
	Coefficient	P-value	Coefficient	P-value	
Perform below ability (never/rarely)	1.365	0.108	1.286	0.174	
Homework (never/rarely)	-1.130	0.312	0.196	0.851	
Absent (never/rarely)	-0.797	0.392	2.517	0.011	
Lazy (never/rarely)	1.128	0.306	-0.801	0.511	
Inattentive (never/rarely)	-1.906	0.131	0.180	0.865	
Disinterested (never/rarely)	1.983	0.112	-1.008	0.364	
Noisy (never/rarely)	2.836	0.046	n/a	n/a	
Mother complete primary	2.230	0.134	-0.486	0.726	
Mother some secondary	3.152	0.077	1.185	0.456	
Mother complete secondary	3.787	0.054	1.372	0.435	
Mother some tertiary	7.350	0.021	5.146	0.064	
Father complete primary	2.720	0.18	3.555	0.062	
Father some secondary	-0.294	0.847	-0.154	0.914	
Father complete secondary	0.227	0.912	-1.369	0.448	
Father some tertiary	2.232	0.254	2.089	0.22	
One durable good	-3.039	0.131	0.953	0.614	
Two durable goods	-4.721	0.012	-3.139	0.07	
Three durable goods	-2.572	0.17	-0.143	0.932	
Four durable goods	-3.773	0.054	-1.188	0.526	
Five or more durable goods	-3.241	0.087	0.261	0.882	
One news item	0.072	0.948	-0.622	0.523	
Two news items	0.210	0.874	3.254	0.008	
No livestock	-1.629	0.07	-0.687	0.411	
Arsenic	0.511	0.642	0.422	0.676	
Non Muslim	-2.138	0.217	-2.265	0.115	
13 years old	1.382	0.137	0.414	0.613	
14 years old	1.078	0.342	-0.450	0.658	
15 years old or older	-1.382	0.403	0.174	0.905	
Female	-1.019	0.158	-0.881	0.172	
Admission any school	-0.502	0.217	-0.378	0.314	
Madrasah attendance	-0.696	0.113	-1.381	0	
School has electricity	-0.691	0.191	-0.807	0.094	
School has a library	0.277	0.497	-0.133	0.721	
School has a computer	0.813	0.05	0.641	0.091	
Number of observations	262		284		
Adjusted R-squared	0.1181		0.1701		

Table A5: Models for mathematics test scores between schools reflecting Barry's view*

*Models include constant term.

Explanatory variables	Dependent var Mathematics	riable: score	Dependent variable: English score		
	Coefficient	P-value	Coefficient	P-value	
Perform below ability (never/rarely)	0.499	0.388	0.709	0.244	
Homework (never/rarely)	-0.394	0.58	-0.219	0.747	
Absent (never/rarely)	-0.477	0.435	1.495	0.022	
Lazy (never/rarely)	0.340	0.64	0.094	0.902	
Inattentive (never/rarely)	-0.901	0.26	0.006	0.992	
Disinterested (never/rarely)	0.975	0.221	-0.088	0.878	
Noisy (never/rarely)	1.093	0.117	n/a	n/a	
Mother complete primary	2.105	0.164	-0.574	0.677	
Mother some secondary	3.418	0.059	1.716	0.279	
Mother complete secondary	3.862	0.053	1.440	0.408	
Mother some tertiary	7.565	0.019	6.068	0.028	
Father complete primary	2.940	0.15	3.926	0.037	
Father some secondary	-0.557	0.719	-0.378	0.79	
Father complete secondary	-0.102	0.96	-1.682	0.346	
Father some tertiary	2.041	0.299	2.075	0.221	
One durable good	-3.061	0.131	0.368	0.842	
Two durable goods	-4.607	0.015	-3.531	0.038	
Three durable goods	-2.196	0.246	-0.154	0.927	
Four durable goods	-3.510	0.08	-1.219	0.515	
Five or more durable goods	-3.168	0.095	-0.407	0.816	
One news item	0.161	0.886	-0.485	0.615	
Two news items	0.600	0.657	3.727	0.003	
No livestock	-1.585	0.084	-0.572	0.489	
Arsenic	0.824	0.46	0.694	0.489	
Non Muslim	-1.724	0.326	-1.929	0.175	
13 years old	1.092	0.244	0.102	0.902	
14 years old	1.089	0.341	-0.446	0.658	
15 years old or older	-1.454	0.377	-0.467	0.742	
Female	-1.082	0.136	-0.807	0.205	
Admission any school	-0.426	0.3	-0.319	0.388	
Madrasah attendance	-0.375	0.383	-1.391	0	
School has electricity	-0.821	0.124	-1.023	0.03	
School has a library	0.196	0.634	-0.089	0.812	
School has a computer	0.752	0.069	0.745	0.044	
Number of observations	262		284		
Adjusted R-squared	0.1011		0.1828		

Table A6: Models for English test scores between schools reflecting Roemer's view*

*Models include constant term.