Two procedures for assessing inequality of educational opportunities in Brazil

Fábio D. Waltenberg, Universidade Federal Fluminense

Abstract

ENGLISH: In this paper, we discuss advantages and limits of two alternative methodologies which can be used for measuring inequality of educational opportunities, both of which are based on Roemer (1998). The two alternatives reflect the usual opposition between a dominance approach and an approach based on specific indices. We provide illustrations using Brazilian data. The *dominance analysis* reveals a situation of evident inequality of opportunity when types are defined in terms of parental education, while when types are defined in terms of skin color we obtain both inequality of opportunity and (at least weak) equality of opportunity, depending on the types we compare. The *inequality indices* approach shows that, according to the parameters we employ, inequality of opportunity represents 16.1% of overall inequality in Brazil, and we observe large regional variation.

PORTUGUÊS: Neste artigo, discutem-se vantagens e limites de dois métodos alternativos de mensuração de desigualdades de oportunidades educacionais, ambas as quais se baseiam em Roemer (1998). Tais alternativas refletem a oposição usual entre uma abordagens de dominância e aquelas baseadas em índices específicos. Apresentamos ilustrações usando dados brasileiros (do SAEB). A *análise de dominância* revela uma situação de evidente desigualdade de oportunidades quando os tipos são definidos em termos de nível de educação dos pais, enquanto no caso em que se definem por meio da cor da pele, obtêm-se tanto desigualdade de oportunidades como igualdade de oportunidades (fraca), em função dos tipoes que estejam sendo comparados. A abordagem baseada em *índices de desigualdade* mostra que, de acordo com os parâmetros usados, a desigualdade de oportunidades representa ao menos 16.1% da desigualdade total no Brasil, e se observam substanciais variações regionais.

JEL classification: D63, I21, I39.

Key-words/palavras-chave: equality of opportunity, education fairness, inequality measurement, opportunity measurement / igualdade de oportunidades, justiça educacional, medidas de desigualdade, medidas de oportunidade.

ÁREA ANPEC: Área 11 - Economia social e demografia

1 Introduction

In this paper, we discuss different methodologies for measuring and comparing inequality of educational opportunities which are based on, or try to go beyond, Roemer (1998). To do so, we apply a distributional analysis toolbox to subsets of the Brazilian pupils' population (groups or types). We measure and compare *inequality of educational opportunities* across groups or types of pupils, instead of gross inequalities as it is usually done.

Just after this introduction, we provide a methodological discussion (section 2) of two different procedures for measuring inequality of educational opportunity, reviving the usual divide between dominance analysis *versus* specific inequality indices approach. Illustrations of those methodologies are reported in section 3, where we employ a particular Brazil's education dataset (the so-called SAEB). The *dominance analysis* reveals a situation of evident inequality of opportunity when types are defined in terms of parental education, while when types are defined in terms of skin color we obtain both inequality of opportunity and (at least weak) equality of opportunity, depending on the types we compare. The *inequality indices* approach shows that, according to the parameters we employ, inequality of opportunity represents 16.07% of overall inequality in Brazil. We observe large variation across regions, just as in the Italian study we use as a benchmark. Brazil's most opportunity-unequal region (South-East) is 176.67% more unequal than Italy's most opportunity-unequal region (Center-South) and 433.71% more so than Italy's least opportunity-unequal region (North). Section 4 brings final remarks.

2 Methodological issues

2.1 Recipient unit, attribute, and aggregation method

In this chapter, we take pupils as the recipient units, and the absolute scores they obtain as attributes. As for the aggregation method, we would like to use a procedure which, while respecting usual properties, would also be as compatible as possible, with the conception of justice stated by Roemer (1998), described along this section 2.

We discuss two alternative paths for comparing distributions in terms of their inequality of opportunity, both of which do respect standard properties, and are inspired by Roemer's normative criterion. In order to attenuate the inconveniences of the anonymity property, the population is subdivided in groups ("types" in Roemer's terminology), and usual inequality measurement tools are applied to those subgroups.

One of theses paths has been developed in a series of papers: Peragine (1999, 2002, 2004b,a) and especially Checchi and Peragine (2005). The alternative path has been developed by a group of French economists: Pistolesi, Lefranc, and Trannoy (2005); Lefranc, Pistolesi, and Trannoy (2006).

While both strategies are clearly inspired by the work of Roemer, there is a net difference between them, which revives the usual distinction between dominance analysis and inequality indices approach. The procedure developed by the Italian economists is an interesting translation of Roemer's intuition: it is based on the use of inequality indices whose aim is to capture the degree of *unfair* inequalities, providing complete rankings of distributions. The procedure developed by the French economists generalizes, in a certain sense, Roemer's concept of EOp. It provides dominance analysis and, as expected, does not necessarily permit to set up complete rankings, but provides more robust conclusions, at least in some pair-wise comparisons.

We briefly present each of them in the next section. Then we assess some advantages and disadvantages of each methodology.[1]

2.2 Inequality indices

Extending a previous theoretical contribution (Peragine, 2004b) which focused on opportunity inequality dominance criteria, Checchi and Peragine (2005) design a procedure devoted to producing *complete rankings* of *opportunity inequality*. They do so by distinguishing the fraction of inequality which is due to circumstances (unacceptable) and the fraction of inequality which is due to effort (acceptable), and then assessing the absolute magnitude and the relative importance of each of these two kinds of inequality. Sections 1 to 4 in their paper describe in detail their approach and now we offer a brief explanation, emphasizing the points which matter more for us.

The first step is to partition the distribution of an outcome (say, of test scores) according to two criteria: types and "tranches". Belong to a given type all the individuals who have similar circumstances (say, the same family background). Belong to a given tranche all the individuals (from each type) who sit in the same quantile in the distribution of the outcome. To Roemer the absolute level of effort is irrelevant, and it is the effort relative to the individual's type what matters. So all the individuals who belong to a given tranche are assumed to have expended a similar degree of effort and, hence, deserve to attain the same educational level.

In the second step, each observation of scores is replaced by the arithmetic mean score within a given type and tranche. In other words, they replace the original profile of scores by an artificial profile of scores in which all the intra-type-and-tranche inequality fades away. The rationale for doing so is that such within-type-and-tranche inequality is neither associated with differential opportunities, nor with differential effort. Being irrelevant in Roemer's model, it should be canceled out in an empirical application. In the "artificial" distribution generated

[1]It should be acknowledged the existence of another branch of this literature, which tries to split total variance of achievement into circumstance- *versus* choice-related components, based on *regression techniques* (Devooght, 2004; Cognaux and Gignoux, 2005; Bourguignon, Ferreira, and Menéndez, 2006) and relaxing the assumption of separability between effort and circumstances, which is adopted in Roemer's framework. This insightful set of papers will be explored in our future research.

thereby, in accordance with Roemer's assumptions, the scores are exclusively a function of the type (previously set by the researcher) and of the effort (inferred by the researcher from the outcome distribution).

The third step is then to evaluate the overall inequality of the artificially generated distribution, as well as the within-tranches inequality and the within-types inequality. Within-tranches inequality is interpreted as inequality due to differential opportunity, since the outcomes of individuals of different types, but who belong to the same tranche (i.e., assumed to have expended similar relative effort), are being compared. In a world of perfect equality of opportunities, there would be no difference from one type to another, within each tranche, in the outcome level obtained. Within-types inequality, in turn, is interpreted as inequality due to differential effort, since the outcomes of individuals of similar types (i.e., those who have similar circumstances), are being compared.

Concretely, the decomposition of overall inequality of the artificially generated distribution into opportunity-inequality and effort-inequality is analogous to the decomposition of overall inequality into between-groups inequality and within-groups inequality. Groups could be defined both as types or as tranches, and Checchi and Peragine discuss both possibilities.[2] We prefer the tranches approach, which in our view is closer to Roemer's conception of EOp, and we will only refer to such approach hereafter. According to such tranches approach, between-groups inequality should be understood as between-tranches inequality (acceptable); and within-groups inequality should be understood as within-tranches inequality (unacceptable).

In order to decompose inequality, Checchi and Peragine (2005) have employed a subgroup decomposable index, namely an index belonging to the generalized entropy class. They have chosen the mean logarithmic deviation (MLD), also known as 2$^{\text{nd}}$ Theil's coefficient, or as the generalized entropy index, $GE(0)$. They justify the choice of this index because of a property it respects, namely, path-independence, which is useful in the context of their paper.

2.3 Dominance analysis

A contribution by Pistolesi, Lefranc, and Trannoy (2005) provides an alternative framework for comparing distributions in terms of the degree of equality of opportunity they reflect. Such framework places itself in the tradition of dominance analysis. Again, we limit ourselves to a brief explanation here.

Pistolesi, Lefranc, and Trannoy (2005) summarize the opportunity offered to an individual by the distribution of an outcome, s, conditional on her set of circumstances, t, denoted $F(s|t)$. Considering two groups in a society, $F(s|t)$ and $F(s|t\prime)$, if $F(s|t)$ second-order stochas-

[2]The "tranches approach" is stated in their "Definition 2: (...) There is EOp if and only if all those who exerted the same degree of effort have the same chances of achieving the objective, regardless of the type." In our view, this approach is closer to Roemer's conception of EOp than the "types approach" stated in their Definition 1.

tically dominates $F(s|t\prime)$, then the situation of the former group is unambiguously preferred to that of the latter one.[3] In such case, since it is clear that the distribution of one group is preferred to the distribution of another group, such society presents *inequality of opportunity*.

Logically then, equality of opportunity is satisfied only when the conditional distribution of the outcome of one group does not dominate the conditional distribution of the outcome of another group. They say that "defining equality of opportunity as non-dominance with a second order stochastic dominance criterion is equivalent to saying that an individual choosing among these circumstances is unable to rank them."(Pistolesi, Lefranc, and Trannoy, 2005, p.5). However, two kinds of equality of opportunity are possible:

1. *Weak equality of opportunity*: verified whenever it is not the case that one distribution dominates the other one, and provided that the cumulative distributive functions of the two groups (or equivalently, their GL curves) *cross at least once*. The reasoning is the following: given that they cross, it is not clear from an *ex ante* position (i.e., from a 'veil of ignorance') if it is better to belong to type t or to type $t\prime$.
2. *Strong equality of opportunity*: verified whenever it is not the case that one distribution dominates the other one, and provided that the cumulative distributive functions of the two groups (or GL curves) *are identical*. In this case, any *ex ante* observer would have no clue on choosing whether he would prefer to belong to group t or to group $t\prime$.

They claim strong equality of opportunity is a particular case of the weak equality of opportunity. And they also claim such particular case corresponds exactly with Roemer's conception of equality of opportunity.

2.4 Methodological issues: an assessment

The opposition between the two methodologies revives the debate on specific indices *versus* dominance procedure. In fact, the advantages and limits of an approach based on specific indices hold for the 'Italian strategy', while those of the dominance approach hold for the 'French strategy'.

An application of the 'Italian strategy' provides us with numbers (synthetic indices of inequality; fractions of acceptable and of unacceptable inequality), which can be compared within and across regions or countries. That may be a very useful tool to better understand particular patterns of inequality in different places. For example, we are able to compare Brazilian figures we calculate here with Italian figures calculated by Checchi and Peragine (2005). The drawback is

[3]We slightly modify the notation employed by Pistolesi, Lefranc, and Trannoy (2005). While they use x and s for outcomes and circumstances, respectively, we employ s (for scores) and t (for types), respectively.

that in the absence of dominance, we can never be sure whether these comparisons would be robust to another index. And this is precisely the greatest advantage of the French approach.

An interesting feature of the 'Italian approach' is that it is very close in spirit to Roemer's conception of EOp, which is an important reference in the normative literature. While an index such as the one the 'Italian approach' offers is indeed a *specific* index, it can not be considered as a fully *arbitrary* one, given that Roemer's theory/algorithm it expresses is normatively meaningful.

What we consider in the 'Italian approach' to be a departure from Roemer's original EOp conception - the use of the $GE(0)$ index instead of one which would be closer in spirit to a maximin across types, such as $GE(-1)$ - seems to us as being in fact a recommendable change, not only due to the technical reason raised by Checchi and Peragine (2005), but also due to two other reasons. The first one is also somewhat technical. When advising on how to choose an inequality index, Cowell (1995, p.65) suggests to take into account what he calls the "discriminatory power of an inequality measure", an idea which is clearly explained as follows: "if very high inequality aversion is specified, nearly all income distributions that are encountered will register high measured inequality, so that it becomes difficult to say whether one is more unequal than another." An index such as $GE(0)$, without being insensitive with respect to inequality, has potentially more discriminatory power than a more more 'extreme' index such as $GE(-1)$.

There is also a more substantial reason to favor the use of $GE(0)$, following Moreno-Ternero (2005). In his defense of "a more equitable proposal for equality of opportunity", while adopting Roemer's algorithm, he refuses to give absolute priority to the worst-off group at each tranche. In other words, he contests the use of a maximin across types in favor of another normative position, which would reduce the weight attributed to the worst-off group, in order to increase the weight of other groups. The idea underlying his proposal is that, while such groups are not the worst-off, they might also face difficulties in terms of opportunities, and thus deserve some "help" too. The inequality index corresponding to maximin in the GE-class would be $GE(-1)$. An index such as $GE(0)$, in turn, would arguably play the role expected by Moreno-Ternero, since it places less weight on inequality at the bottom of the distribution than $GE(-1)$, but still is sensitive to inequality at the bottom.

One limit of the 'Italian procedure' regards its 'second step', where each observation of scores is replaced by the arithmetic mean score within a given type and tranche, aiming at canceling out irrelevant inequality. Clearly, that is an arbitrary discrete approximation of the original continuous distribution. In their paper, Checchi and Peragine (2005) test the impact of such approximation, by comparing the inequality of the original and of the generated distributions, and they find out that the impact is very small. (We make the same test in the empirical section of this paper, reaching a similar result.) However, it is true that the final figures of the 'Italian procedure' (i.e., those expressing opportunity-inequality) obtained both in their study (and in ours) can be criticized on those grounds. Two paths for future research seem to be possible. Either studying the systematic impacts, if any, of different approximations (through sensitivity analysis, etc.) and then selecting the most appropriate among them; or turning to other approaches, such as the regression-based ones, already mentioned before.

According to the 'French approach', Roemer's definition of EOp is only a particular case of a more general conception of EOp in which an individual choosing *ex ante* among different circumstances is unable to rank them (EOp as non-dominance with a second order stochastic dominance criterion). As we said above, according to them, Roemer's position corresponds with "strong EOp", a situation in which the CDFs of every pair of types are identical. For two CDFs to be identical, the frequency distributions from which the CDFs derive must also be identical. We do not interpret Roemer's EOp as requiring identical frequency distributions across every pair of types, at least whenever we take into account Roemer's second-best compromise (or averaging formula), cf. Roemer (1998). Indeed, generally, for Roemer's EOp to hold, outcomes must be identical *only in each tranche* and not (necessarily) along the whole distribution. Then, in the averaging formula used to define the actual EOp allocation rule, the same weight is given to each tranche (or quantile), such that after the EOp algorithm has been implemented (i.e., a reallocation of resources has moved society into a new situation) the frequency distributions (and thus the CDFs) of the outcome for each type needs not (necessarily) be identical. So, while it is true that Roemer's EOp without the averaging formula would correspond with 'strong EOp', Roemer's EOp plus the averaging formula does not. In any case, whatever the label given to Roemer's EOp, the 'French approach' is a very welcome dominance procedure for fairness analysis.

The two approaches discussed here are attractive for different reasons and useful for different objectives. In our view, they can be employed in complementary ways, given that each will provide contrasted perspectives on the same issue - of measuring unfair inequalities of test scores. In the remaining of this chapter, we employ each of them in turn.

3 Inequality of educational opportunities in Brazil

3.1 Data, definitions, and descriptive statistics

We use here the SAEB dataset, produced by INEP, a federal autarchy subordinated to the Ministry of Education[4]. We report national and regional descriptive statistics regarding only 2001, math, and 8$^{\text{th}}$ grade,

[4]Due to space limitation, we refrain from describing in detail the SAEB dataset here. Detailed information can be found in the INEP webpage: http://www.inep.gov.br/.

except when we calculate SF, because in such case we need to use data regarding the three grades.

We need to condition test scores upon individuals' circumstances (*types* in Roemer's terminology). One of the types on which we condition scores is pupils' parental education (the highest education level among parents). Let us mention a few reasons why we use such information to define types. First of all, because pupils' parental education seems to appropriately play the role of a "relevant circumstance". Starting with the so-called "Coleman report" (Coleman et al., 1966), a vast literature in the economics of education has established that the influence of variables related to socio-economic status (SES) on achievement is considerable. Pupil's parental education typically is positively correlated to other SES variables (wealth, occupational status, type of neighborhood where family lives in, and so on), so it can be viewed as a variable summarizing SES. The role of SES as a determinant of achievement is particularly strong in Brazil, as have shown studies based, for example, on SAEB data (Albernaz, Ferreira, and Franco, 2002). Finally, while not talking about performance in test scores, but about years of schooling, Cognaux and Gignoux (2005) usefully remind us that Brazil has one of the lowest intergenerational educational mobility, reinforcing the evidence that parental education plays an important in shaping educational opportunities.

Another requirement for a variable to define a type is that it is not under control of the individuals, nor subject to manipulations by them. While the true information on pupil's parental education is not under control of the individuals (in the sense that it is exogenous to a particular individual), the value which is reported can be manipulated by pupils, since often it is a self-reported variable. However, we do not see why pupils would adopt any kind of strategic behavior in this respect when answering a questionnaire such as the one accompanying the SAEB exams. Yet, it should be acknowledged that such variable is subject to misreporting, a problem self-reported variables are typically subject to.

Parental education is usually available in test scores datasets, as well as in other datasets, which ensures comparability. For example, in the specific case of the empirical application we undertake here, we are able to compare our results to those coming from a similar study made using Italian data. In the future, we could also reproduce the exercise using PISA, and other national and international datasets, producing thereby comparable results, which would be helpful to put in perspective the ones we provide in this chapter.

Pupils' parental education is thus a good candidate for a "relevant circumstance", since it has a great influence on the outcome of pupils, it is often available, and for practical matters it is not under control of pupils.

In Table 1 are reported some descriptive statistics on scores at both the national and the regional level. We observe in Panel A the descriptive statistics concerning the pooled sample. Panel B shows the statistics for the sub-sample of observations for which information on

Table 1: Checking impact of missing information.

Panel A: Pooled sample

Region	Mean Score	Std. Dev.	Obs.	Freq.
North	231.86	42.44	7,972	188,469
North-East	228.79	46.13	20,166	800,674
South-East	249.72	51.83	8,672	1,379,759
South	255.34	45.96	6,251	397,108
Mid-West	244.83	45.76	7,239	236,262
Brazil	243.38	49.62	50,300	3,002,272

Panel B: Subsample with missing information on parents' education

Region	Mean Score	Std. Dev.	Obs.	Freq.
North	219.88	36.72	559	14,623
North-East	217.48	41.27	1,671	71,163
South-East	230.4	46.00	636	102,581
South	241.54	39.52	383	25,945
Mid-West	231.59	39.08	499	19,022
Brazil	227.13	43.49	3,748	233,334

parental education is missing. There is missing information concerning 3,748 observations out of a total of 50,300, which represents 7.45% of the total (or 7.77% of the weighted observations). Another important feature we can read in this table is that the average score obtained by the pooled sample (243.38) is higher than the average score obtained by the sub-sample (227.13, which corresponds to 93.32% of the pooled sample's average score).[5] It is not surprising that pupils who are not able to report their parents' education obtain lower scores than those who are able to do so. From now on, we only use the 46,552 observations for which we have information on parents' education, but the slight bias mentioned here should be acknowledged.

In Table 2 we can observe descriptive statistics of scores, by region and by type of pupil. The national average score for the valid observations is 244.75, with standard deviation 49.86. The $46,552$ valid observations expand to $2,768,938$ pupils with the sample weights.[6] The most populated region is by far the South-East ($1,277,178$ weighted observations), more than seven times larger than the least populated region (North, with $173,846$ weighted observations).

As parents' education increase, average scores increase too, both in the national level (ranging from 213.11 to 286.52) and inside each region. The ranking of regions in terms of average scores is as follows: S $\succ$ SE $\succ$ MW $\succ$ N $\succ$ NE. This overall ranking varies slightly according to the type we turn to. For example, among kids with highly-educated parents (college), the ranking is: SE $\succ$ S $\succ$ MW $\succ$ NE $\succ$ N, while the ranking among kids with poorly educated parents (lacking formal education) is: MW $\succ$ SE $\succ$ S $\succ$ N $\succ$ NE.

3.2 Specific inequality indices

In this section are reported results obtained regarding Brazilian educational data, following the procedure exposed in section 2.2 and previously applied by Checchi and Peragine (2005) to Italian data.

3.2.1 Preliminary checks

The first step is to check how scores vary according to quantiles (deciles here) and to pupils' types (their parents' education here), which is shown at the country level in Table 3. For a given type, we observe similar frequencies across deciles.[7] We also observe that, at any given decile, scores increase with the type (as we move from the top to the

[5]The same pattern is observed for regions' statistics. The fraction of missing observations ranges from 6.13% in the South to 8.29% in the North-East; while the ratio between sub-sample average scores and pooled-sample average scores ranges from 92.26% in the South-East to 95.06% in the North-East.

[6]In fact, the $46,552$ valid observations expand to $2,768,938.40$, but we omit the decimals in this table and in the remaining ones.

[7]*Frequencies* are *similar*, but not equal, from the construction of the per-type deciles (in terms of weighted frequencies, and not in terms of observations) and the discrete nature of scores (with continuous data, we would have identical frequencies).

Table 2: Descriptive statistics related to pupil's mother's education.

Means, Standard Deviations, Frequencies and Number of Observations of Score

Parents' Education	**Region**					**Brazil**
	N	NE	SE	S	MW	
No Formal	216.68	207.40	217.91	216.94	224.89	213.11
	36.41	35.12	34.09	41.12	39.26	36.03
	7,516	61,937	39,905	8,106	8,845	126,309
	253	1,170	188	76	182	1,869
Lower Prim.	223.51	216.95	234.15	243.03	229.88	229.29
	37.49	37.79	41.71	40.68	38.20	40.98
	42,967	234,858	358,861	99,650	56,536	792,873
	1,537	4,600	1,429	1,014	1,169	9,749
Upper Prim.	227.07	224.60	238.58	248.11	237.43	235.85
	39.45	39.79	44.03	41.20	39.55	42.72
	45,317	182,362	362,977	110,756	62,813	764,224
	1,668	3,895	1,710	1,290	1,531	10,094
Secondary	238.48	241.83	259.68	262.75	253.49	253.28
	41.84	46.60	50.91	44.64	44.05	48.58
	54,104	173,069	295,437	93,295	54,752	670,657
	2,337	5,081	2,127	1,561	1,728	12,834
College	253.00	273.05	294.93	289.17	281.71	286.52
	51.43	57.11	54.84	47.83	51.69	55.15
	23,942	77,286	219,998	59,356	34,294	414,876
	1,618	3,749	2,582	1,927	2,130	12,006
Total	232.87	229.90	251.28	256.31	245.99	244.75
	42.73	46.43	51.97	46.23	46.12	49.86
	173,846	729,512	1,277,178	371,163	217,240	2,768,938
	7,413	18,495	8,036	5,868	6,740	46,552

Table 3: Scores according to types and deciles. (Table continues in the next page...)

Means, Standard Deviations, Frequencies and Number of Observations of Score

Parents' Education	**Decile**					**Total**
	1	2	3	4	5	
	158.34	176.95	186.97	195.71	203.95	213.11
No Formal	8.13	4.02	2.96	2.28	2.35	36.03
	12,648	12,652	12,766	12,461	12,637	126,309
	187	170	164	148	175	1,869
	165.19	185.39	198.16	210.58	221.74	229.29
Lower	9.50	4.01	3.43	3.70	2.95	40.98
Primary	79,350	79,718	78,916	79,224	79,246	792,873
	1,025	948	1,022	1,13	972	9,749
	166.59	189.35	203.78	216.81	228.20	235.85
Upper	10.44	4.41	4.07	3.48	3.16	42.72
Primary	76,595	76,534	76,140	76,543	77,293	764,224
	944	967	1,089	996	948	10,094
	174.95	200.45	217.38	231.40	244.45	253.28
Secondary	11.73	5.84	4.20	3.75	3.92	48.58
	67,072	67,235	67,149	66,934	67,017	670,657
	1,084	1,196	1,110	1,121	1,269	12,834
	188.42	223.98	246.35	265.79	282.52	286.52
College	16.07	7.25	5.95	5.34	4.23	55.15
	41,499	42,653	40,336	41,497	41,552	414,876
	861	925	1,055	1,232	1,214	12,006
	171.10	195.63	210.95	224.95	237.30	244.75
Total	14.23	14.57	17.30	19.78	21.75	49.86
	277,164	278,792	275,307	276,659	277,745	2,768,938
	4,101	4,206	4,44	4,627	4,578	46,552

Table 3: (...Continued from previous page)

Means, Standard Deviations, Frequencies and Number of Observations of Score

Parents' Education	**Decile**					**Total**
	6	7	8	9	10	
	213.51	223.85	236.06	254.11	282.71	213.11
No Formal	3.08	3.33	4.12	6.13	17.33	36.03
	12,644	12,627	13,182	12,075	12,615	126,309
	186	206	196	216	221	1,869
	232.88	244.45	256.73	271.62	306.51	229.29
Lower	3.40	3.23	3.74	5.37	21.11	40.98
Primary	79,330	79,342	79,174	79,787	78,786	792,873
	1,012	907	881	932	920	9,749
	240.05	252.90	267.00	282.06	312.07	235.85
Upper	3.51	4.04	4.16	4.73	19.44	42.72
Primary	75,432	76,463	76,480	76,414	76,330	764,224
	1,079	1,076	992	870	1,133	10,094
	258.22	271.68	286.91	305.65	341.97	253.28
Secondary	3.93	3.66	5.09	6.42	20.28	48.58
	66,988	67,152	67,094	67,015	67,001	670,657
	1,328	1,266	1,376	1,369	1,715	12,834
	297.19	312.07	326.88	345.81	376.92	286.52
College	3.73	4.28	4.44	6.94	13.29	55.15
	41,442	41,542	41,407	41,466	41,482	414,876
	1,134	1,408	1,468	1,441	1,268	12,006
	249.77	262.58	276.35	293.10	326.14	244.75
Total	23.16	24.28	25.26	26.85	32.71	49.86
	275,837	277,126	277,337	276,758	276,215	2,768,938
	4,739	4,863	4,913	4,828	5,257	46,552

bottom of the table). We confirm thereby a pattern which was expected.[8]

Another check we have undertaken, following Checchi and Peragine (2005), is a comparison of the original distribution and the artificially generated distribution (the one which cancels out the within-tranche-and-type inequality). The results of this check (not reported here) show the transformation does not affect inequality very much, since the ratio between the inequality - according to $GE(0)$ - of the two distributions is always around 1, both at the national level, and for each region.

3.2.2 Main results

The main results of this section are reported in Table 4, in which we observe the decomposition of overall inequality into opportunity-inequality and effort-inequality.

Based on the parameters we have used, which express a very parsimonious definition of circumstances, we immediately observe two important rankings - columns A and A/C - which equally rank Brazil's regions: N (smaller inequality and smaller fraction) $\succ$ S $\succ$ MW $\succ$ NE $\succ$ SE (large inequality and larger fraction).

In terms of *magnitude of inequality of opportunity*, the result for Brazil's fairest region (North, 0.0009449) is comparable with Italy's results. Inequality of opportunity in that region is 33.27% larger than that of Italy's fairest region (Northern Italy, 0.000709) and 30.91% smaller than that of Italy's unfairest region (Center-Southern Italy, 0.0013677). Thus, those results place Northern Brazil in an intermediary situation between the two Italian regions.

However, when we turn to the second fairest region in Brazil (South, 0.0020801), inequality of opportunity is already much more substantial: 52.08% more than Italy's *un*fairest region. Brazil's unfairest region shows a level of inequality of opportunity (South-East, 0.0037840) which is 176.67% larger than that of Italy's unfairest region (Center-Southern Italy, 0.0013677) and 433.71% larger than that of Italy's fairest region (Northern Italy, 0.000709). So, with the exception of the North, unfairness is much greater in Brazilian regions than in the Italian regions. Considering that only a small fraction of Brazilian pupils study in the Northern region (6.28% of the final weighted sample we used) the overall picture is one of considerable inequality of opportunity in Brazil.

The *fraction of inequality which is due to differential opportunity* amounts to 16.07% of overall inequality in Brazil, and we observe substantial variation across regions. The North is the least unequal one with 5.66%, while the South-East is the most unequal one with 18.06% of overall inequality due to inequality of opportunity.

[8] We have also made the same calculations for each region. The same pattern observed at the national level is verified, with a few violations of the monotonicity of scores with types: (i) region=SE, decile=1, types= 1, 2, 3; (ii) region=S, decile=1, types= 2, 3; (iii) region=MW, decile=6, types= 1,2.

Table 4: Decomposing inequality into opportunity and effort inequality

	Inequality			Percentage
	...of opportunity	...of effort	Overall	Opportunity/Overall
	(A)	(B)	(C)	(A/C)
Brazil	0.0032685	0.0170664	0.0203350	16.07%
North	0.0009449	0.0157431	0.0166879	5.66%
North-East	0.0029423	0.0164729	0.0194152	15.15%
South-East	0.0037840	0.0171662	0.0209502	18.06%
South	0.0020801	0.0145051	0.0165852	12.54%
Mid-West	0.0024914	0.0147000	0.0171914	14.49%

Inequality of effort is comparable across regions (column C), ranging from 0.0145051 (South) to 0.0171662 (South-East), with a national level of 0.0170664. These levels of effort inequality are not considerably different from the range found by Checchi and Peragine (2005), namely, 0.0126289 in Northern Italy and 0.0174451 in Center-Southern Italy. So the difference in the magnitude of scores inequality between Italy and Brazil is essentially due to opportunity-inequality, and not to effort-inequality.

3.3 Dominance analysis

In this section, using Brazilian educational data, we apply the procedure exposed in section 2.3, and which has been previously applied by Pistolesi, Lefranc, and Trannoy (2005) to international data on incomes.

According to Pistolesi, Lefranc, and Trannoy (2005) the opportunities offered to an individual of a given type are summarized by the CDF of her type. By plotting CDFs of different types we can verify, for every pair of CDFs, whether there is inequality of opportunity (curves do not cross), weak equality of opportunity (curves cross) or strong equality of opportunity (curves are identical).

We recall that, adapting to our purposes the notation defined by Pistolesi, Lefranc, and Trannoy (2005), we set the conditional scores function as $F(s|t)$, where s stands for individual pupils' score, as usual, and T stands for type. Type is defined as five categories of pupils' parental education. Since we are using exactly the same data as in the previous section, refer to Tables 2 and 3 for descriptive statistics.[9]

[9] In the CDFs graphics, we use the label *isced* for parental education level, and we have: *isced*1: no formal education; *isced*2: lower primary education; *isced*3: upper primary education; *isced*4: secondary education; *isced*5: college education.

3.3.1 Main results

Figure 1 depicts the CDFs for the 5 types of pupils in Brazil. We can observe that the higher the level of parents' education (*isced*5) the better is the performance of pupils. There is no crossing, which means that we are in a situation of *inequality of opportunity*, for any pair of types we look at. We can observe, though, that the gap across types is not uniform. For example, the difference from the curve associated with parents having higher education (*isced*5) and that of parents having high-school certificate (*isced*4) is larger than the difference between the curves associated with *isced*4 and *isced*3. The largest differences are to be found between the two highest performing types of pupils (*isced*5 and *isced*4) and between the two lowest performing types of pupils (*isced*2 and *isced*1). The same pattern is reproduced almost exactly at each of the 5 macro-regions (regional graphs not reported here).

Figure 1: CDFs, with parents' education (isced) defining types

We still have to check whether these results resist statistical inference tests (in our plans for future research), especially whether the types *isced2* and *isced3* are not in a situation of strong equality of opportunity.

3.3.2 An alternative definition of types

We have repeated the exercise using an alternative definition of types: pupil's skin color. Which definition of types - in terms of skin color, of parental education, or of a combination of the two - represents a more legitimate definition of types is an open question, which has recently given rise to an intense debate in Brazil concerning the legitimacy of affirmative action policies for higher education. An affirmative action policy for higher education would certainly benefit relatively well-off non-white individuals (those who would have surpassed uncountable obstacles to reach the more advanced levels of the Brazilian educational system). While it may be legitimate to grant access to college to non-whites in order to ensure the emergence of diversity among the elite - with possible positive effects, for example, in terms of enhanced motivation for non-white children, and so on -, if the objective is to actually improve opportunities for a larger fraction of non-whites, affirmative action policies would have to be implemented in much earlier stages of the schooling system, possibly at primary school. The objective of the exercise here is simply to contribute to such debate, diagnosing whether inequality of opportunity in terms of skin color can be identified in test scores data, that is, when individuals are still at school.

The variable we employ here originates from a choice made by pupils for defining their skin color themselves, among 5 possible categories: white, mixed, black, Asian, native-Brazilian. Descriptive statistics regarding this variable in its relation with scores are reported in

Table 5: Descriptive statistics related to pupils' skin color

Means, Standard Deviations, Frequencies and Number of Observations of Score

Race	**Region**					**Brazil**
	N	NE	SE	S	MW	
White	234.74	231.22	260.57	259.67	250.36	252.93
	45.36	49.47	53.35	46.88	48.64	52.07
	57,976	260,924	667,992	278,067	99,225	1,364,183
	2,742	6,887	4,482	4,668	3,494	22,273
Mixed	230.81	227.91	239.56	244.53	242.54	235.55
	39.99	44.14	46.73	41.48	43.41	45.13
	92,400	367,518	476,853	78,521	89,581	1,104,874
	3,733	8,947	2,642	964	2,476	18,762
Black	225.94	221.09	225.11	240.60	230.13	225.24
	44.29	40.83	41.33	44.25	38.33	41.56
	12,542	83,869	112,775	16,933	19,292	245,411
	456	1,875	668	253	457	3,709
Asian	242.13	241.46	259.29	255.77	247.69	251.68
	43.26	49.31	58.50	43.93	43.50	53.11
	13,847	50,244	86,226	15,363	18,263	183,943
	637	1,500	592	233	550	3,512
Native	226.34	229.88	246.39	251.64	239.82	238.23
	38.30	43.46	42.60	32.74	43.99	42.88
	7,371	22,368	25,769	4,019	6,324	65,852
	271	611	206	81	173	1,342
Total	232.39	229.20	249.99	255.59	245.17	243.75
	42.39	46.17	51.83	45.99	45.72	49.62
	184,137	784,923	1,369,614	392,903	232,686	2,964,262
	7,839	19,820	8,590	6,199	7,150	49,598

Table 5. There is less missing information on pupils' self-reported skin color (702) than on parents' education (3748), which turns to be an advantage of defining types on the basis of skin color over defining them on the basis of parental education.[10] One characteristic of this definition is that the groups (types) created this way are too different in terms of size: while whites and mixed together amount to 82.74% of the total population, native-Brazilians are only 2.71% of the total, and blacks and Asians amount to less than 8% each. In some regions, the presence of some types is negligible. For example, the sum of blacks (4.08%), Asians (3.76%) and native-Brazilians (1.31%) amounts to less than 10% in the South, while in that same region 3/4 of pupils report themselves as being white.

An interesting result with this new definition of types is that clearcut situations of both kinds are found: equality and inequality of opportunity. In fact, we find: (i) equality of opportunity (at least in the weak sense) between whites and Asians, and between mixed and native-Brazilians, and (ii) inequality of opportunity between blacks and any other group, and between the pair mixed/native and the pair white/Asians.

Figure 2: CDFs, with self-reported skin color defining types

Although such results should be verified by statistical inference, they lead us to provisionally conclude that there is inequality of opportunity for achievement in terms of skin color in Brazil.

4 Final remarks

In the methodological sections of this chapter, we discuss advantages and limits of two alternative methodologies which can be used for measuring and comparing inequality of educational opportunities, both of which are based on, or try to go beyond, Roemer (1998).

The *dominance analysis* reveals a situation of evident inequality of opportunity when types are defined in terms of parental education, while when types are defined in terms of skin color we obtain both inequality of opportunity and (at least weak) equality of opportunity, depending on the types we compare. The *specific indices* approach show that inequality of opportunity represents 16.07% of overall inequality in Brazil, and we observe large variation across regions. Brazil's most opportunity-unequal region (South-East) is 176.67% more unequal than Italy's most opportunity-unequal region (Center-South) and 433.71% more so than Italy's least opportunity-unequal region (North).

Various, alternative, specifications can be tested and a series of improvements can be implemented to both applications (we have mentioned a few in the text). It is also possible to deepen our analysis focusing on particular sub-national units, in order for our results to be

[10] We have no information on the reliability of those two self-reported variables.

more useful for policy use. We plan to accomplish those tasks in the close future.

References

ALBERNAZ, A., F. H. G. FERREIRA, AND C. FRANCO (2002): "Qualidade e eqüidade no ensino fundamental brasileiro," *Pesquisa e Planejamento Econômico*, 32(3).

BOURGUIGNON, F., F. H. G. FERREIRA, AND M. MENÉNDEZ (2006): "Inequality of opportuny in Brazil," mimeo.

CHECCHI, D., AND V. PERAGINE (2005): "Regional disparities and inequality of opportunity: the case of Italy," IZA Discussion Paper, n° 1874.

COGNAUX, D., AND J. GIGNOUX (2005): "Earnings inequalities and educational mobility in Brazil over two decades," IAI Discussion Papers n° 121.

COWELL, F. A. (1995): *Measuring inequality.* Prentice Hall, Harvester, Wheatsheaf, 2nd edn.

DEVOOGHT, K. (2004): "To each the same and to each his own. A proposal to measure responsiblity-sensitive income inequality," mimeo.

LEFRANC, A., N. PISTOLESI, AND A. TRANNOY (2006): "Une réduction de l'inégalité des chances dans l'obtention du revenu salarial en France?," mimeo.

MORENO-TERNERO, J. D. (2005): "A more equitable proposal for equality of opportunity," Unpublished manuscript. Available at http://merlin.fae.ua.es/juande/papers/EOp.pdf.

PERAGINE, V. (1999): "The distribution and redistribution of opportunity," *Journal of economic surveys*, 13(1), 37–69.

——— (2002): "Opportunity egalitarianism and income inequality," *Mathematical Social Sciences*, 44, 45–64.

——— (2004a): "Measuring and implementing equality of opportunity for income," *Social Choice and Welfare*, 22, 187–210.

——— (2004b): "Ranking income distributions according to equality of opportunity," *Journal of Economic Inequality*, 2, 11–30.

PISTOLESI, N., A. LEFRANC, AND A. TRANNOY (2005): "Inequality of Opportunity vs. Inequality of Outcomes: Are Western Societies All Alike?," ISER Working Paper Series, n. 2005-15, University of Essex, UK.

ROEMER, J. (1998): *Equality of opportunity.* Harvard University Press, Cambridge, MA.

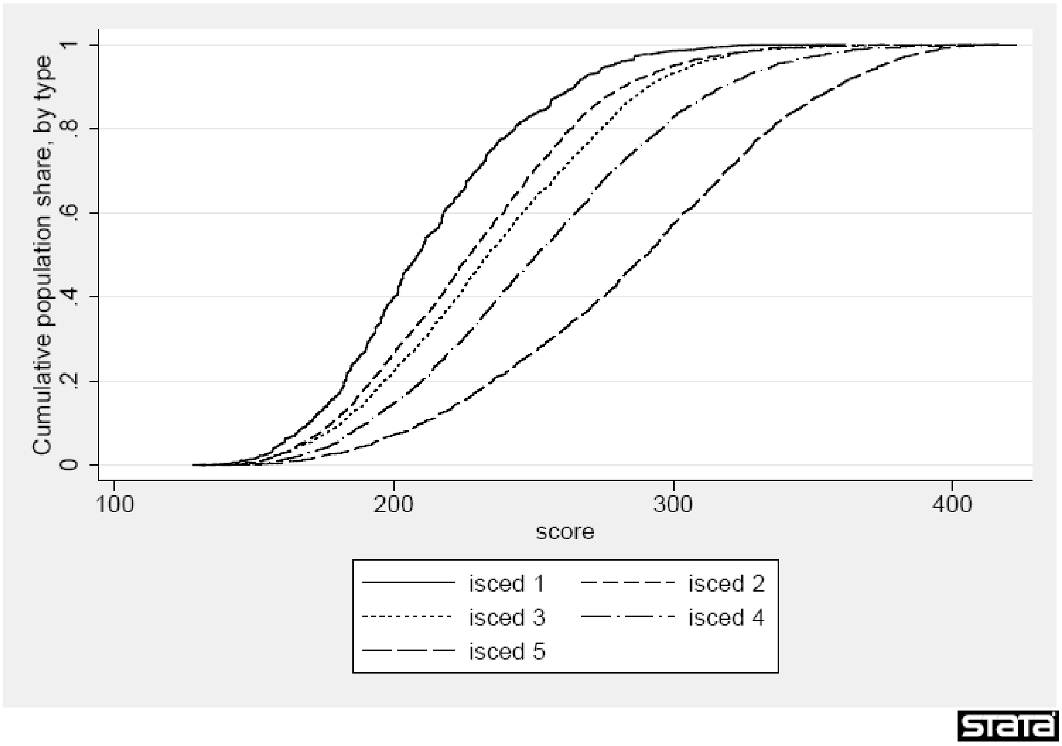

Figure 1: CDFs, with parents' education (isced) defining types

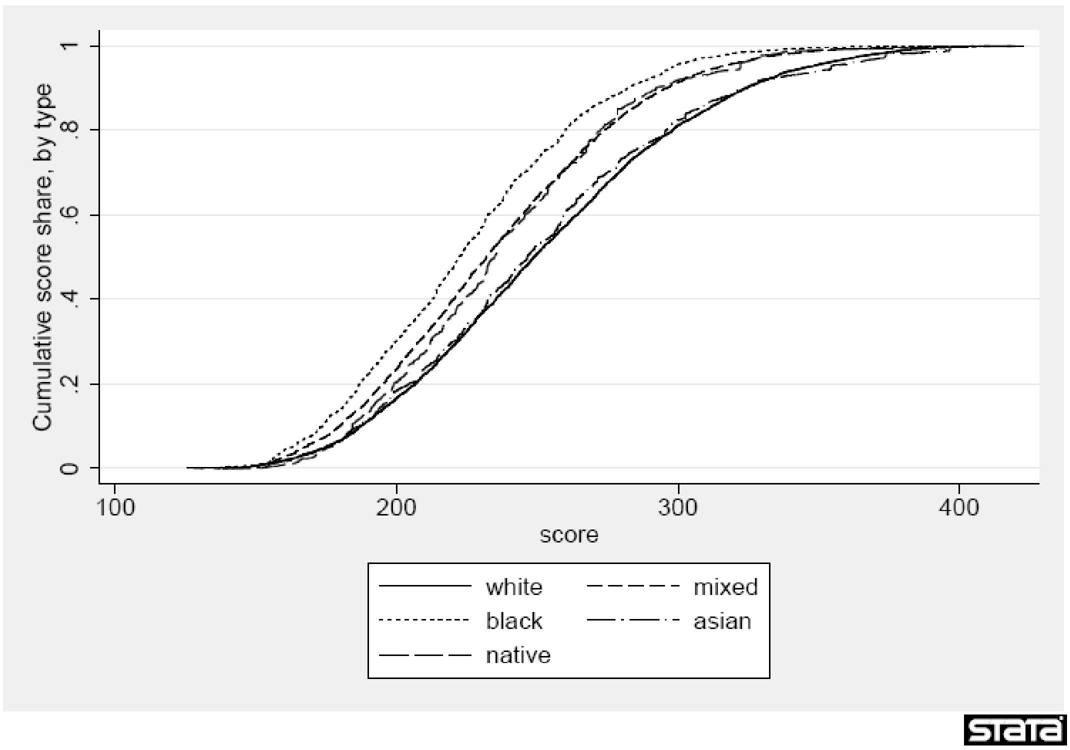

Figure 2: CDFs, with self-reported skin color defining types