Equality of Opportunity with qualitative variables: an application to education in Europe

Carmen Herrero
University of Alicante & Ivie

Ildefonso Mendez
University of Murcia

Antonio Villar
University Pablo de Olavide, OECD & Ivie

Abstract:

This paper deals with the study of Equality of Opportunity when agents' achievements are expressed in terms of ordered categorical data. We provide a cardinal evaluation function defined on the distribution of achievements of the different types (groups that gather agents with similar circumstances). Following the work of Herrero & Villar (2013), we apply here this evaluation criterion to the analysis of equality of opportunity in education regarding some European countries.

Keywords: Equality of opportunity, categorical data, relative frequencies, domination probabilities, European countries, education.

1 Introduction

Equality of Opportunity (EOp) is nowadays one of the most prominent concepts of distributive justice. The key idea behind this concept is that the concern about inequality should not focus on the equality of outcomes but rather on the existence of a common playing field for all people in society. From this perspective agents' outcomes can be regarded as deriving from two different sources, usually referred to as *effort* and *opportunity*. Effort alludes to the result of people's decisions whereas opportunity relates to the agents' external circumstances. A fair society is one in which final outcomes do not depend much on the agents' external circumstances, that is, a society in which all people share similar opportunities. In that society outcome differences are basically determined by the agents' preferences and effort and not by aspects that are beyond their control and responsibility (see Arneson, 1989; Cohen, 1989; Roemer, 1993, 1998).

From a methodological perspective this type of approach implies that we cannot make direct comparisons of outcomes corresponding to people who have different circumstances. In order to deal with this difficulty, the analysis of equality of opportunity usually follows three steps. First, agents are classified into different *types*, according to their circumstances. Second, outcome differences within types are considered as ethically irrelevant, as they correspond to differences in effort. And third, a measure of between types differences is devised in order to approximate the relative situation of those types and/or the degree of dependence of outcomes on circumstances.

The equality of opportunity principle is important and elusive. It is important because of its philosophical underpinnings and its policy implications. In particular, it follows from this approach that people who are relatively disadvantaged due to external circumstances deserve some kind of compensation, whereas those outcome differences that derive from their effort are perfectly acceptable.[1] It is elusive because it is not always easy to decide which variables can be regarded as circumstances and

[1] There is a wide spectrum of views with respect to what is required for equality of opportunity, from the non-discrimination view to the view that social provision should compensate for all forms of disadvantage. Common to all these views is that individuals are accountable, to some extent, for the achievement of the advantage in question, whether this refers to health, education, income, utility or welfare. The issue of responsibility has become prominent in some of the recent developments within the areas of political philosophy and welfare economics (e.g. Fleurbaey (2008) and the literature cited therein).

which variables are related to the non-observable notion of effort. Think for instance on the possible influence of luck or whether agents are fully responsible for their preferences, including time discount and risk attitudes. Or, in a dynamic context, when outcomes derived from past effort may be regarded as future circumstances. There is not yet a well-established consensus on the proper way of addressing those problems.[2]

John Roemer (1988, 1993) suggested a way of measuring inequality of opportunity by comparing the quantile distribution of achievements of agents of different types. He considers that two individuals have exerted a similar level of effort whenever they are located at the same quantile of their type distribution. In this way the possible dependence on the distribution of effort on the type is taken into account. The differences in the level of the variable that defines those quantiles for the types provide quantitative measures of the impact of circumstances on outcomes. Then he proposes an infinite inequality aversion measure to provide an estimate of the inequality of opportunity.[3] Key in this approach is the way of defining the types that gather people with similar circumstances.

A different way of capturing the notion of opportunity is in terms of "opportunity sets". An individual's opportunity set consists of all possible outcomes that she might achieve by exerting different levels of effort, given her circumstances. Here again we find the difficulty of dealing with the non-observable "opportunity sets", mostly when dealing with empirical analysis. There is a number of proposals in the literature that approximate the individuals' opportunity sets by means of some nonparametric estimate. Van de gaer (1993) uses an average of the achievements of all people in each type, in the income distribution case. Lefranc et al (2008) measure the value of the opportunity set by the surface under the Lorenz curve of the income distribution of the individual's type. Once this has been established, inequality of opportunity is captured by using some specific inequality measure (e.g. the Gini index).

Here we follow the opportunity set approach to measure inequality of opportunity. Assuming a common support for the possible realizations of all types, we approximate the opportunity set of an individual by the distribution of achievements of those agents of the same type. The idea is that such a distribution tells us what is

[2] There are, nonetheless, several papers that provide estimates of inequality of opportunity in particular contexts, as Bourguignon et al., 2007, Checchi & Peragine, 2010, or Almas et al., 2011.

[3] See Moreno-Ternero (2007) for an alternative way of measuring inequality of opportunity within Roemer's framework.

feasible for a representative agent of a given type and how likely is for that agent to get each feasible outcome. The key informational input is, therefore, the distribution of realizations of agents across types. From this perspective, comparing opportunities amounts to comparing distributions of achievements across types. When society is made of many agents the relative frequencies of the different achievements provide a sufficient description of the probability distributions. In order to compare those distributions we apply some of the ideas introduced in Herrero & Villar (2013, 2014a), which permit dealing with categorical data (assuming that categories are ordered, as it happens in many relevant situations).

Rather than focussing on an overall measure of equality of opportunity, we shall present a vector of evaluations that inform about the relative opportunities of the different types (i.e. each type gets a score that tells about the relative richness of its opportunity set). Such a vector corresponds to the eigenvector of a suitably constructed matrix that describes the likelihood of pairwise dominance of the different types. This measure can be regarded as an extension of some of the ideas used in the analysis of discrimination and segregation, that usually apply just to the two-groups case (e.g. Echenique & Fryer (2005), Chakravarty & Silber (2007), Frankel & Volij (2011)). It has also some points in common with other measures based upon spectral properties of certain matrices describing networks relations. This is the case, in particular, of the analysis of intellectual influence of publications, analysed in Palacios-Huerta & Volij (2004).

Our interpretation of equality of opportunity is in line with the idea of capability sets proposed in Sen (1985), which has been used in many different contexts (e.g. Herrero (1996), Herrero & Pinto (2008), or the UN Human Development Index). The possibility of treating inequalities when the statistical information is given in categorical data appears in Allison & Foster (2004) in relation with self-reported health status (see the applications in Abu-Naga & Yalcin (2008) and that of Dutta & Foster (2011), regarding the measurement of inequalities in happiness). Yalonetzky (2012) also deals with inequalities with ordinal variables. Different approaches to the equality of opportunity concept appear in Thomson (1987), (1994).

The paper is organised as follows. Section 2 presents the formal model and proposes a way of measuring relative equality of opportunity. Section 3 analyses this proposal from an axiomatic viewpoint in terms of the properties it fulfils. Section 4 provides an empirical illustration of the working of our measure, making use of the SHARE dataset. We analyse the equality of opportunity in education, by comparing the access to education in eight European countries, as a function of the parental education, in two different population cohorts. Section 5 gathers a few final comments.

2 The Model

Consider a society made out of $N = \{1, 2, \ldots, n\}$ individuals. We want to evaluate the situation of that society, regarding a particular aspect (education, health, labour, access to income, etc.), from an Equality of Opportunity viewpoint. The achievements of the population with respect to such an aspect are described by a finite set of ***levels of achievement***, $L = \{1, 2, \ldots, s\}$, ordered from best to worst. We can think of those levels as categories (e.g. health statuses, educational levels, professional positions) or as intervals of a continuous variable (e.g. income brackets, age intervals).[4]

Suppose that we have identified a collection of variables that allow classifying the individuals in N into τ different ***types***, $T = \{1, 2, \ldots, \tau\}$, depending on their circumstances. That is, we assume that opportunity differences induce a partition of individuals into types, so that all agents within a type share the same circumstances (i.e. they have the same opportunity set). Differences in the achievements *within* a type correspond, therefore, to differences in effort between people with similar traits. Differences *between* types, on the contrary, reflect the different opportunities in society.

We shall address the evaluation of opportunities problem following the *veil of ignorance* approach. That is, we consider the situation faced by an agent who tries to evaluate the different opportunities in society without knowing who she will be within each possible type. This amounts to finding a ranking of the different types valid for any transitive and monotone preference relation. The evaluation problem implies, therefore, comparing the relative goodness of the distribution of achievements within each type, trying to maximize the likelihood of higher levels of achievement.

[4] Continuous variables can also be accommodated easily in the model, even though we shall not deal with this case here, in order to emphasize the applicability of this approach to categorical variables. For a discussion of this aspect see Herrero & Villar (2014a). This type of problem has also been dealt with, following a different approach, in Herrero & Villar (2014b)

Let a_{ir}, $i = 1, \ldots, \tau$, $r = 1, \ldots, s$, be the proportion of people of type *i* with level of achievement *r*. This information can be collected in a matrix **A** whose *i*th row describes the distribution of achievements of the agents of type *i*, in terms of relative frequencies. An ***evaluation problem***, or simply a *problem*, can thus be characterized by a set of individuals, *N*, a set of types, *T*, a set of levels of achievement, *L*, and the matrix **A** of relative frequencies that describes the distribution of achievements by types. We shall denote a problem by $P = (N, T, L, \mathbf{A})$ and call Ω the set of all evaluation problems. An ***opportunity evaluation function***, *F*, is a mapping such that, for each evaluation problem *P*, it associates a vector $\mathbf{v} = F(P)$ that provides a summary measure of the relative opportunities of the different types ($\mathbf{v} \in R_+^t$, where t is the number of types in the problem). Note that our notion of evaluation function is vector valued, that is, we aim at providing an evaluation of each type relative to the others rather than a real-valued summary measure (we shall return to this point later on).

Comparing the relative opportunities between two types, from the veil of ignorance approach, amounts to comparing the relative chances of having a better result. Let p_{ij} be the probability that an agent of type *i* chosen at random exhibit a higher level of achievement that an agent of type *j*, randomly chosen. As the levels of achievements are ordered, this probability can be simply obtained as follows:

$$p_{ij} = a_{i1}(a_{j2} + \ldots + a_{js}) + a_{i2}(a_{j3} + \ldots + a_{js}) + \ldots + a_{i(s-1)}a_{js}$$

Similarly, let p_{ji} be the probability that a representative member of type *j* exhibits a higher level of achievement that a representative agent of type *i*. And let $e_{ij} = 1 - p_{ji}$ the probability of an agent of type *i* getting the same level of achievement than an agent of type *j*.

Given an evaluation problem *P*, we say that type *i* ***dominates*** type *j* in a binary comparison, when $p_{ij} > p_{ji}$. In pairwise comparisons, the quotient p_{ij} / p_{ji} informs us about such a relative opportunity. If $p_{ij} / p_{ji} > 1$, people of type *i* have better opportunity than people of type *j*, and viceversa. When $p_{ij} / p_{ji} = 1$, both types have the same opportunity.[5]

[5] Lieberson (1976) in a similar vein introduces the *Index of Net Difference*, $ND(i, j)$, as the absolute difference between p_{ij} and p_{ji}, to inform about inequalities between two groups. If $ND(i, j) = 0$, then $p_{ij} = p_{ji}$ that is, it is equally likely, given an individual chosen at random in any of the two groups that the individual in i is at a better position than the individual in j than the other way around. The other extreme case is when $ND(i, j) = 1$; which happens whenever all individuals in one of the groups are at better positions than those in the other group. Intermediate positions provide with values of $ND(i, j)$ between 0 and 1.

This notion of domination is an intuitive and sensible way of comparing relative opportunities out of the outcome distributions of pairs of types. If type *i* dominates type *j*, then the opportunity set of the first type can be regarded as better than the opportunity set of the second, as the representative agent gets on average better outcomes. So the circumstances that define type *i* are preferable to those of type *j*. Note that we have not only a ranking of opportunity sets but also a cardinal measure of their relative opportunity (the size of the ratio p_{ij} / p_{ji}).

Things are not so easy when it comes to evaluating more than two types. Unfortunately, the application of this evaluation criterion to more than two types fails to be transitive, and thus we cannot rank the types, even though we may know the relative evaluation of any pair of them. The example in Table 1 illustrates the problem. It describes the distribution of the population of three different types, 1, 2 and 3, into four levels of achievements, I, II, III and IV.

Table 1

Levels of achievement	I	II	III	IV
Shares of type 1	0.1	0.3	0.6	0
Shares of type 2	0.5	0	0	0.5
Shares of type 3	0	0.7	0.2	0.1

Computing the corresponding domination probabilities yields the following results: (i) $p_{12} = 0.5$, $p_{21} = 0.45$, which implies that type 1 dominates type 2. (ii) $p_{23} = 0.5$, $p_{32} = 0.45$, which implies that type 2 dominates type 3. And (iii) $p_{31} = 0.42$, $p_{13} = 0.25$, which implies that type 3 dominates type 1, thus creating a cycle.

As a consequence, when society consists of more than two types we have to depart from the mere pairwise comparisons, i.e, when valuing the relative opportunities of a type w.r.t some others, all have to be taken into account simultaneously.

In order to keep the *veil of ignorance* approach, as well as the likelihood of having better opportunities, we define an evaluation protocol that extends the idea of domination probabilities. Suppose we have an evaluation problem involving a society with t types and we want to assess the relative likelihood of achieving higher levels. To do so we apply repeatedly a tournament procedure defined as follows. We start by selecting arbitrarily a pair of types, *i* and *j*, say, and take one as the incumbent and the other one as the defiant. Suppose we choose *i* as the incumbent (eve though the choice is immaterial). Next, pick randomly one agent of each of those types. We then compare the levels of achievement of those two agents. If the individual of type *i* belongs to a higher or equal level of achievement than the individual of type *j*, then type *j* is dismissed while type *i* will be confronted with another type *k* randomly selected.[6] Two new individuals are now chosen at random, one from the incumbent type *i* and one from the defiant type *k,* and are confronted as before. Once more, the type whose representative agent exhibits a higher level of achievement remains, the other one is dismissed, and a new round starts. By repeating infinitely many times this process, we find that the probability that *i* remains in the contest is given by:

$$\frac{\sum_{j \neq i}\left(p_{ij} + \varepsilon_{ij}\right)}{t - 1}$$

(the sum of the weak domination probabilities of *i* over the remaining types, times the probability of *i* being chosen at random). Similarly, the probability that type *i* be beaten by type *j* is given by $p_{ji} / (t - 1)$ (the probability that type *j* dominates type *i*, times the probability of *j* being chosen).

This protocol defines, therefore, a Markov process whose stochastic matrix is:

$$M(P) = \frac{1}{t-1}\begin{pmatrix} \sum_{j\neq 1}\left(p_{1j}+\varepsilon_{1j}\right) & p_{12} & \dots & p_{1t} \\ p_{21} & \sum_{j\neq 2}\left(p_{2j}+\varepsilon_{2j}\right) & \dots & p_{2t} \\ \dots & \dots & \dots & \dots \\ p_{t1} & p_{t2} & \dots & \sum_{j\neq t}\left(p_{tj}+\varepsilon_{tj}\right) \end{pmatrix}$$

This matrix has a positive dominant eigenvector, $\mathbf{v}^* = M(P)\mathbf{v}^*$, which corresponds to a stationary distribution.[7] This vector **v*** indicates the proportion of time that (in the long run) each type will be selected in the above protocol. Those values provide a measure of the relative goodness of the distribution of outcomes within each type and thus capture the quality of the opportunity set of each type. We can, therefore, identify our opportunity evaluation function with this stationary distribution, that is: For each problem *P*, $F(P) = \mathbf{v}^*$, where **v*** is the dominant

[6] The distinction between incumbent and defiant serves just the purpose of deciding the type that remains in the tournament when both agents exhibit the same level of achievement.

[7] This vector always exists for this type of matrix and will be unique and strictly positive provided matrix *M*(*P*) is irreducible. See for instance Berman & Plemmons (1994).

eigenvector of matrix $M(P)$.

Note that the ith component of our evaluation function is given by:

$$F_i(P) = \frac{\sum_{j \neq i} p_{ij} F_j(P)}{\sum_{j \neq i} p_{ji}}, \quad i = 1, 2, \ldots, t$$

That is, the ratio between the weighted sum of the domination probabilities of i with respect to all other types, with weights equal to their corresponding evaluations, and the sum of the probabilities that i be dominated by some other type.

Remark: *This is, precisely, the notion of* ***worth*** *introduced in Herrero & Villar (2013). This vector can be immediately calculated by means of a freely accessible algorithm designed for that purpose (see* http://www.ivie.es/valoracion/index.php).

Note that, for the case of two types, we have $\frac{F_i(P)}{F_j(P)} = \frac{p_{ij}}{p_{ji}}$ so that this protocol is actually an extension of the probability of domination criterion. In two-type problems those probabilities fully determine the evaluation (except for the choice of units).

* * *

The evaluation function F provides a measure of relative equality of opportunity of the different types in society. Each type receives a score that reflects the relative goodness of the distribution of opportunities in enjoys. This is the key feature of the proposal. The dispersion of those scores tells us about the degree of inequality of opportunity in society. So one may consider defining a synthetic measure of the society's equality of opportunity following the normative approach used in the evaluation of income distribution. That is, inequality of opportunity can be measured by an index:

$$I_{Opp}(P) = 1 - \frac{v^e(\mathbf{v}*)}{m(\mathbf{v}*)}$$

where $v^e(\mathbf{v}*)$ is the egalitarian equivalent worth, that is the number that plugged into the implicit welfare evaluation function W would yield the same value that the one associated to the actual worth vector. Formally:

$$W\left[\underbrace{v^e(\mathbf{v}^*), \ldots, v^e(\mathbf{v}^*)}_{\tau \text{ times}}\right] = W(\mathbf{v}^*)$$

which, under a suitable scaling assumption yields: $W(\mathbf{v}^*) = m(\mathbf{v}^*)\left[1 - I_{Opp}(\mathbf{v}^*)\right]$.

Setting the inequality measure amounts to defining the implicit welfare function, and viceversa. We shall propose here a specific way of measuring the inequality of opportunity: the Atkison's inequality measure for the unitary value of the inequality aversion parameter. This results into an egalitarian equivalent worth given by the geometric mean of the components of vector **v***, that we denote by $\tilde{\mu}(\mathbf{v}^*)$. And, setting $m(\mathbf{v}^*) = 1$, we can take $I_{Opp}(\mathbf{v}^*) = 1 - \tilde{\mu}(\mathbf{v}^*)$ as the appropriate measure of inequality of opportunity in society and $W(\mathbf{v}^*) = \tilde{\mu}(\mathbf{v}^*)$ as the corresponding equality of opportunity measure.

3 Properties of the evaluation function

In order to provide additional support to our opportunity evaluation function, we follow the *axiomatic approach,* by showing that it satisfies a number of interesting properties that reinforce its operational and normative appeal.[8]

Let *P* = (*N*, *T*, *L*, **A**) be an evaluation problem involving *n* agents, t types, and *s* levels of achievement. We present now different properties that the evaluation function described in the former section satisfies. The first one worth mentioning is that of *anonymity*, which says that the evaluation only depends on the characteristics of the types and not on other aspects such as labels or names. Formally:

- **Anonymity:** *For any problem P, for any permutation* $\Pi : T \to T$ *and for all* $i \in T$, *it happens that* $F_{\Pi(i)}(P) = \Pi\left(F_i(P)\right)$

A second standard property has to do with the situation in which for all pairs of groups, i, j it happens that $p_{ij} = p_{ji}$, that is, in all pairwise comparisons there is no advantage of any group over the other. If this is the case, then all groups should be

[8] About the relevance and interest of the axiomatic approach, see Thomson (2001).

given identical value.

- **Symmetry:** *For any problem P such that for all* $i, j \in T$, $p_{ij} = p_{ji}$ *, we have that* $F_i(P) = F_j(P)$*, for all* $i, j \in T$.

Consider now the particular case of evaluation problems involving only two groups. The next three properties refer to that case and are closely related. We start by introducing a *monotonicity* property that describes the behaviour of the evaluation function when there is an improvement in one of the groups. Let $P = (N, \{i, j\}, L, \mathbf{A})$ be a two-group problem and suppose that some members of group *i* improve their situation by moving to a higher position (from *s* to *r*, say, with $r < s$, $s, r \in L$). Let $P' = (N, \{i', j\}, L, \mathbf{A})$ denote the new situation and call the change from *i* to *i'* an *unambiguous improving.* Then, the relative valuation of group *i* with respect to group *j* should be higher in problem *P'* than it was in the original problem. Formally:

- **Monotonicity:** *Let* $P = (N, \{i, j\}, L, \mathbf{A})$ *be a two-type problem, and let* $P' = (N, \{i', j\}, L, \mathbf{A})$ *be a problem that results from an unambiguous improving of group* i. *Then,* $\frac{F_i(P)}{F_j(P)} < \frac{F_i(P')}{F_j(P')}$.

The next property, *stochastic dominance*, can be regarded as an implication of the monotonicity and anonymity properties presented above. It relates the valuation of two groups of the same size, when one of them stochastically dominates the other one.

- **Stochastic Dominance:** *Let* $P = (N, \{i, j\}, L, \mathbf{A})$ *be a two-group problem and suppose that the following holds:*

$$a_{i1} \geq a_{j1},$$
$$a_{i1} + a_{i2} \geq a_{j1} + a_{j2},$$
$$\cdots$$
$$a_{i1} + a_{i2} + \cdots + a_{i,s-1} \geq a_{j1} + a_{j2} + \cdots + a_{j,s-1}$$

with some strict inequality. Then $\frac{F_i(P)}{F_j(P)} > 1$.

The dominance relation between groups *i* and *j* before can be interpreted as *i* being an unambiguous improving of *j* (either in one or several steps). Thus, combining monotonicity with anonymity, we get the stochastic dominance property. Also note that, in this context, stochastic dominance implies dominance in pair-wise comparisons.

Consider finally, the idea of providing a specific form of the relative evaluation of pair of groups, whenever we face a two-group problem. *Reciprocity* establishes that, in those problems involving only two groups, the relative valuation of the groups should coincide with the ratio between the number of times a group beats the other.[9] Formally:

- **Reciprocity:** *For any problem* $P = (N, \{i, j\}, L, \mathbf{A})$, *we have:*

$$\frac{F_i(P)}{F_j(P)} = \frac{p_{ij}}{p_{ji}}$$

Previous property implies both monotonicity and stochastic dominance.

The properties analysed so far involve a *fixed* number of types. We now introduce a property that, instead, relates problems with different number of types. This is a property in which we require a sort of *bilateral consistency* in the way the solution behaves when considering a series of two-type associated problems.[10]

In order to deal with the case of three or more types we introduce now the notion of ***opportunity advantage***. Given a problem *P* and an evaluation function *F*, the opportunity advantage of type *i*, $a_i(P,F)$, is defined as the weighted sum of the probabilities that this type dominates the others, where the weights are given the evaluations of those other types. Formally:

$$a_i(P,F) = \sum_{j \neq i} p_{ij} F_j(P)$$

It is easy to see that when there are only two types, type *i* exhibits higher opportunity advantage than type *j* if and only if type *i* dominates type *j*. This is not true when there are more than two types.

---

[9]A similar property appears in Slutzki & Volij (2006)

[10] On the relevance and ideas about bilateral consistency in different settings, see Hokari & Thomson (2008).

A simple way of treating the case of more than two types is by applying a certain requirement of *consistency.* We do so by establishing a link between the case of more than two types, $\tau > 2$, and several related two-type problems.

For a given problem P = (*N*, *T*, *L*, **A**), with $\tau > 2$, let $P^{(i)} = (N, \{i, N_{-i}\}, L, \mathbf{A}^{(i)})$ denote the two-type problem resulting from merging all types different from *i* into a single one. This is done by taking $a_{-ij} = \frac{1}{\tau - 1} \sum_{k \neq i} a_{kj}$. Consistency requires that the ratio between the opportunity evaluation for type *i* in the two problems coincide with the ratio of the corresponding opportunity advantages. Formally,

- **Consistency:** *Let P be a problem. Then, for all* $i \in T$,

$$\frac{F_i(P)}{F_i\left(P^{(i)}\right)} = \frac{a_i(P,F)}{a_i\left(P^{(i)},F\right)}$$

The idea of consistency has been widely used in the literature of bargaining problems, coalitional games or bankruptcy situations, as a way of dealing with variable population problems. For an excellent survey on consistency, see Thomson (2011), (2012). The idea behind this principle in this context is that changing the way of conforming the types should not affect the essence of the evaluation.

Among the mentioned properties, the last two, reciprocity and consistency suffice to determine the evaluation function. Me omit the easy proof.

Claim: *An evaluation function F satisfies reciprocity and consistency, if and only if, for any given problem P, all* $i \in T$, *we have:*

$$F_i(P) = \frac{\sum_{j \neq i} p_{ij} F_j(P)}{\sum_{j \neq i} p_{ji}}$$

4 Intergenerational equality of opportunity in Education in Europe

Education is probably one of the fields in which the equality of opportunity approach fits better, as educational performance is clearly correlated to the children's environment (see for instance OECD (2013)). It is also a field in which we find a number of studies addressing specifically this point, taking into account the influence of family characteristics. The idea is that an educational system exhibits a higher degree of equality of opportunity the more independent are the students' outcomes on their family features. An element that has been systematically singularized in order to assess the degree of equality of opportunity is that of the education of the parents (typically the highest educational level achieved by the mother or the father). See for instance Peragine & Serlenga (2008), Lefranc, Pistolesi & Trannoy (2009), Checchi & Peragine(2010), Herrero, Méndez & Villar (2014), Villar (2014).

We follow here this line of analysis by linking equality of opportunity to the independence on the students' achievements on their parents' education. Data on the education level of parents and their children come from the Survey of Health, Ageing and Retirement in Europe (SHARE), a cross-national survey with a panel design representing the population of individuals aged over 50 years in some European countries.[11] SHARE reports the education level of the two parents even if one of them died, thus attenuating sample selection bias due to heterogeneity in mortality rates both by educational levels and across countries. This is particularly relevant in our application since we focus on the father's education because of his higher attachment to the labour market relative to that of the mother, especially for older parents. SHARE also informs on the children's level of education for families of at most four children.[12] However, this limitation is likely to have a reduced impact in our analysis since only 6% of the European parents in the sample have more than four children.

We consider the SHARE dataset corresponding to the first wave relative to the

[11] Further information about the SHARE database is provided in the following link: http://www.share-project.org/

[12] When there are more than four children in the family, the children for which the education level is asked are chosen by sorting them in ascending order by age and geographical distance to the family and picking the first four children. That is, younger children living closer to the family are prioritized.

year 2004.[13] Here we consider three different levels for the father's education (1: primary education or less, 2: intermediate education, and 3: university education), and four different levels for the children's education (1: primary education or less, 2: inferior intermediate education, 3: superior intermediate education, and 4: university education). The education categories in SHARE are defined according to the International Standard Classification of Education (ISCED) and, thus, they are harmonized across countries. We focus on children aged over 25 years in order to minimize the share of children enrolled in formal education. Groups with similar circumstances thus correspond to "children" whose parents have a similar education level.

We concentrate on eight European countries: Sweden, the Netherlands, Spain, Italy, France, Denmark, Greece and Belgium. Germany is excluded from the analysis because there are almost no parents in the lowest educational category. That is so because Germany is the only sample country in which the lowest education degree requires more than seven years of education and, thus, that degree is classified as education level 2 in ISCED. Austria is also excluded for its reduced sample size.

We take two different cohorts for the fathers: those born between 1914-1929, aged 75 to 90 years old in 2004, and those born between 1940-1954, aged 50 to 65 years old in 2004. This allows us to compare the situation of the children of these two cohorts and analyse if the effect of the parents' education level has changed over time.

The descriptive statistics appear in Table 1. For each country and each cohort we have the distribution of children in each of the four education levels considered for children and the three education levels of the fathers. Therefore, each cell (i, j) of the table, other than those corresponding to "All" either in rows or columns, describes -for a given country- the percentage of children that have fathers with educational level j -for a given cohort- who have achieved educational level i. If we consider, for the sake of illustration, the cell (1, 1) of the table for the younger cohort of Spanish fathers, we observe that 19.6% of the children coming from the less educated fathers have the same education level as their fathers. In the rows and columns labelled "All" the percentages are calculated with respect to the total number of persons in the sample

[13] We exploited the panel design of SHARE to minimize nonresponse issues. For households in the first two waves of SHARE, we checked whether the information on the predetermined variables used in the analysis and not reported in 2004 was reported in 2005. That could be the case if, for example, the respondent was not the same in the two interviews. In that case, the information is imputed to the corresponding variables in the first wave.

from that country and that cohort, in brackets. For example, cell (5, 1) of the same table tells us that 45.8% of the Spanish younger fathers had educational level 1, whereas cell (1, 4) says that 9.9% of the Spanish children coming from the younger cohort of fathers had only achieved level 1.

Table 1

		Cohort 1940-1954				Cohort 1914-1929			
	Child's	Fathers' education level				Fathers' education level			
Country	education	1	2	3	All	1	2	3	All
Spain	1	19,6	1,1	3,7	9,9	39,3	1,0	0,0	32,6
	2	29,6	17,9	2,8	21,4	32,3	40,0	2,6	31,9
	3	30,9	42,6	29,9	35,7	17,2	23,8	31,6	18,7
	4	19,8	38,5	63,6	33,0	11,3	35,2	65,8	16,8
	All	45,8	41,9	12,3	100 (869)	82,7	12,7	4,6	100 (825)
Italy	1	3,8	0,6	0,0	2,0	16,6	1,0	0,0	12,3
	2	48,2	19,0	3,1	30,8	43,2	17,3	2,3	35,4
	3	40,3	57,3	33,8	48,4	35,0	52,8	23,3	38,4
	4	7,7	23,1	63,1	18,8	5,3	28,9	74,4	13,9
	All	43,8	50,1	6,1	100 (1070)	72,9	22,2	4,9	100 (886)
Greece	1	6,0	0,9	0,0	2,8	26,4	2,4	2,8	21,2
	2	10,5	4,4	0,5	6,1	16,2	4,7	2,8	13,6
	3	61,2	57,2	36,5	54,9	40,1	43,5	38,9	40,6
	4	22,3	37,4	63,1	36,2	17,3	49,4	55,6	24,6
	All	40,5	40,3	19,1	100 (1061)	78,1	18,1	3,8	100 (940)
Sweden	1	0,8	0,3	0,7	0,5	5,5	2,6	0,7	4,0
	2	13,2	8,3	2,4	8,5	28,8	17,2	9,5	15,8
	3	66,7	59,6	41,7	57,9	43,2	48,5	33,3	30,1
	4	19,3	31,8	55,2	33,1	22,5	31,8	56,5	21,0
	All	27,8	51,8	20,4	100 (1414)	57,3	27,8	14,9	100 (986)
Denmark	1	5,8	0,6	0,7	1,1	7,5	1,9	0,8	3,2
	2	20,4	11,9	6,7	10,7	14,7	9,1	3,3	9,1
	3	53,4	52,3	33,9	45,7	47,6	45,6	24,2	40,5
	4	20,4	35,3	58,7	42,6	30,2	43,4	71,7	47,2
	All	8,4	55,1	36,5	100 (1226)	27,7	46,0	26,3	100 (911)
France	1	14,8	4,6	1,9	7,2	26,0	5,2	4,1	17,9
	2	8,9	5,4	1,6	5,7	10,4	11,5	8,2	10,4
	3	46,0	40,5	20,6	38,1	41,4	39,4	23,8	38,6
	4	30,3	49,4	75,9	49,1	22,1	43,9	63,9	33,1
	All	30,5	48,8	20,7	100 (1547)	61,7	25,3	13,0	100 (1134)
Netherlands	1	8,9	1,4	0,7	2,0	8,7	1,3	1,2	3,1
	2	45,2	25,3	8,8	22,4	47,7	26,6	9,6	28,1

	3	35,0	42,2	25,1	36,2	27,7	40,4	25,7	34,3
	4	10,8	31,1	65,4	39,4	15,9	31,7	63,5	34,4
	All	10,8	58,8	30,4	100 (1453)	24,1	55,3	20,6	100 (810)
Belgium	1	5,9	2,2	0,2	2,3	13,7	1,3	2,2	6,7
	2	15,5	10,6	2,2	9,1	27,0	10,8	3,1	16,2
	3	51,0	41,8	23,9	38,4	40,5	40,3	23,2	37,1
	4	27,6	45,4	73,6	50,1	18,8	47,6	71,4	40,0
	All	18,0	53,9	28,0	100 (1608)	42,0	39,1	19,0	100 (1182)

A simple visual inspection of Table 1 shows that Southern European countries like Greece, Italy and Spain stand out for the low education of their population of fathers. These countries show both the highest percentages of less educated fathers and the lowest percentages of better-educated fathers in the two cohorts. For fathers aged 75 to 90 years old in 2004, the share of those less educated reaches 83% in Spain. Close to the Spanish record are Greece and Italy with 78% and 73% of the fathers in that cohort reporting primary education or less, respectively. These numbers sharply contrast with those for Central and Northern European countries, where the average percentages of less educated fathers in that cohort are both close to 43%. Conversely, while the share of better-educated fathers does not exceed 5% in Southern Europe, it reaches 18% and 20% in Central and Northern European countries, respectively.

The average educational attainment of Southern European fathers improves substantially when we move from the older to the younger cohort of fathers. In absolute terms the larger reductions correspond to those countries, as illustrated by the cases of Spain and Greece with some 37 percentage points of reduction in their shares of low educated parents. In relative terms all countries exhibit large improvements ranging from a reduction of 70% for Denmark to 40% in Italy concerning that variable.

The average educational level of children also improved from the older to the younger cohort of fathers. Here again we find that the higher absolute reductions correspond to Spain and Greece, where the share of less educated children lowered by 23 and 18 percentage points, respectively, whereas the highest relative reductions are those of Sweden (88%), Greece (87%) and Italy (84%), with a minimum of 35% for the Netherlands.

Next, we perform two different exercises using the evaluation protocol presented in sections 2 and 3. First, we evaluate the relative opportunities of children in the two cohorts, for each of the eight countries. We use bootstrap to test whether the difference between the eigenvectors of the two cohorts is statistically significant or not.[14] The results appear in Table 2.

Table 2

	Spain		Italy		Greece		France	
Fathers'	Father's age		Father's age		Father's age		Father's age	
education level	50-65	75-90	50-65	75-90	50-65	75-90	50-65	75-90
1	0.283***	0,123	0,127	0,065	0,310	0,246	0,310	0,369
2	0,783	0,535	0,455	0,385	0.685**	1,223	0.672**	0,917
3	1,934	2,342	2,418	2,550	2.006**	1,531	2.018**	1,714
Summary indicators								
Inequality of Opp.	0,246	0,464	0,481	0,600	0,248	0,228	0,251	0,166
3 over 2	2,47	4,38	5,31	6,63	2,93	1,25	3,01	1,87
3 over 1	6,82	19,06	19,02	39,05	6,48	6,22	6,51	4,65

	Sweden		Denmark		The Netherlands		Belgium	
Fathers'	Father's age		Father's age		Father's age		Father's age	
education level	50-65	75-90	50-65	75-90	50-65	75-90	50-65	75-90
1	0,403	0,453	0,392	0,370	0,214	0,268	0,309	0,226
2	0,729	0,757	0,804	0,659	0,629	0,676	0.616*	0,801
3	1,867	1,790	1,805	1,970	2,157	2,056	2,074	1,973
Summary indicators								
Inequality of Opp.	0,181	0,150	0,172	0,216	0,338	0,281	0,266	0,291
3 over 2	2,56	2,36	2,25	2,99	3,43	3,04	3,37	2,46
3 over 1	4,63	3,95	4,61	5,32	10,07	7,68	6,70	8,73

The eigenvectors are normalized so that the sum of its components equals the number of components. That is, values above 1 indicate that that group has relative

[14] We implement bootstrap hypothesis testing as follows. For each country, let n^a_{pc} indicate the number of children with education level c whose father has education level p and belongs to cohort a, for a = 1, 2. For each country and for each combination of c and p, we merge the samples of the two cohorts of fathers into one sample of $n^1_{pc} + n^2_{pc}$ observations. We draw a bootstrap sample of $n^1_{pc} + n^2_{pc}$ observations with replacement from the merged sample and we assign the first n^1_{pc} observations to the first cohort of fathers. We then calculate the eigenvectors of the two cohorts and we compute the difference between them. We repeat these steps 2000 times. The p-value is then estimated as the number of times the difference between the eigenvectors coming from bootstrap samples exceeds that observed in the original sample, divided over the number of repetitions.

opportunities over the mean, while components below one are to be interpreted as being opportunity deprived. We have included in the table, as the first summary indicator, the measure of inequality of opportunity consisting of $I_{Opp}(.) = 1 - \tilde{\mu}(.)$, where $\tilde{\mu}(.)$ stands for the geometric mean of the components of the corresponding eigenvector.

We observe what is well documented in the literature: the persistence of better opportunities for the children coming form better educated parents, in all countries and in the two cohorts. The components are systematically increasing in the education level of the fathers.

The relative advantage of children coming from better-educated fathers is highest in Italy for both cohorts of fathers. Italy is also the country in which the education opportunities of children coming from less educated fathers are worse in the two cohorts. That is, the highest inequality of opportunity is found in Italy. Conversely, the difference in relative education opportunities of children coming from less and better educated fathers becomes lowest in Sweden and Denmark, the countries with the lowest inequality of opportunity. In Spain, the situation was closer to that of Italy in the older cohort, but it has markedly changed ever since. The eigenvector component for children coming from less educated Spanish younger fathers approximately doubles that in the older cohort. As a result, the ratio between the relative education opportunities of children coming from better and less educated Spanish fathers lowered by a third of its value in the older cohort and the dispersion in the vector components lowered by almost one half. No other European country has experienced such a decrease in inequality of opportunity.[15]

We also can analyse the statistical significance of the evolution of the relative opportunities in the two cohorts. In Sweden, the Netherlands, Italy and Denmark we cannot reject the hypothesis of the samples of the two cohorts to be extracted from the same distribution. That is, we cannot conclude that there has been any change in the relative access to education for children coming from differently educated fathers. The only clear-cut case is that of Spain, in which there has been a significant change in favour of relative EOp. Here, children coming from less educated parents improve relatively their education levels. This also can be observed in the standard deviation of

[15] The percentage change in the inequality of opportunity index for those countries goes as follows: Spain -47%, Italy -20%, Greece +9%, France +51%, Sweden +21%, Denmark -21%, The Netherlands +20%, and Belgium -8%.

the vector components, which diminished in about half of its value. That is the only case in which the change in the vector components is significantly different from zero at the 1% significance level. For the rest of the countries where significant changes are observed, the situation is not so clear. In France and in Greece it seems that there is a relative deterioration of the opportunities of children coming from parents with intermediate education, and an improvement from those coming from better educated parents. This fact comes induces an increment of the inequality of opportunity index, indicating a deterioration of EOp over time. In Belgium, we also obtain a significant difference pointing to the deterioration of the opportunities of children coming from fathers with intermediate education. However, contrary to France and Greece, that difference is only marginally significant for Belgium.

In a second exercise, we compare the relative situation of the individuals in the different countries. The results are in Table 3.

Table 3

	Father's education level					
	1		2		3	
	Father's age		Father's age		Father's age	
Country	50-65	75-90	50-65	75-90	50-65	75-90
Spain	0.580***	0,385	0.923**	0,598	0,871	1,019
Italy	0.452*	0,557	0,582	0,731	0,902	1,490
Greece	1.373***	0,841	1.190**	1,568	0,935	0,670
Sweden	1,455	1,391	0.940*	0,763	0,671	0,659
Denmark	1.049***	1,964	0.961**	1,228	0.724***	1,269
France	1.258**	1,023	1.487***	1,090	1.564***	0,829
Netherlands	0.455***	0,797	0,652	0,669	0,903	0,836
Belgium	1.378***	1,042	1,266	1,351	1,429	1,227
Inequality of Opp.	0,10	0,10	0,04	0,06	0,04	0,04
Max/Min	3,22	5,10	2,55	2,62	2,33	2,26

In this case, the distribution of relative opportunities has changed for the two cohorts, at any level of the fathers' education. The inequality of opportunity for children with parents less educated diminishes, which means that opportunities of those children are now more equally distributed than before across countries. On the other hand, it happens that still in the last cohort, the chances for education of people whose parents had the lowest education level in Sweden, Greece, Belgium and France, are significantly higher than in Italy, the Netherlands or Spain. This means that even though the second cohort fares better in terms of equality of opportunity, there are still relevant differences.

When parents have at least intermediate education, only France, Belgium and Greece keep having significantly higher opportunities. We also find a reduction in the dispersion of the vector components for children coming from fathers with intermediate education.

France is the only country that has significantly improved its relative position at any level of the fathers' education when comparing the older and the younger cohorts of fathers. In line with the findings in Table 2, the improvement is largest for French children coming from the most educated parents. According to the results in Table 3, Spain has also significantly improved its position relative to other European countries in the education opportunities of children coming from less and intermediately educated fathers. On the contrary, Denmark significantly lowered its relative advantage at any level of the father's education levels, particularly so for the children of less educated fathers.

The convergence across European countries in relative opportunities of children coming from less and intermediately educated fathers, reflected in the lower dispersions of the components of the vectors, is driven by changes in the relative positions of Denmark, Greece and Spain. While Denmark lowered its relative advantage, Spain became less opportunity unequal and Greece converged to the mean from above in the case of children of parents with intermediate education, and from below in the case of children of less educated parents.

Also note that this convergence between countries is compatible with the persistence of much higher inequality of opportunity values within countries.

5 Final Remarks

Equality of opportunity can be regarded as a methodological approach that tries to capture the degree of fairness in a society, with respect to a given dimension, in terms of the dependence on the agents' outcomes of their own free decisions and their external circumstances. The higher (resp. lower) the correlation between outcomes and effort (resp. external circumstances), the higher the degree of fairness.

Measuring equality of opportunity requires making a number of compromises on the way of defining effort and opportunity in an operational way, on the one hand, and on the way of aggregating the information on the agents of different types, on the other.

Here we have proposed a strategy of analysis based in the following principles. First, agents are classified into groups that gather people with similar external circumstances. Second, the outcome distribution of the agents of the different groups is regarded as an expression of their opportunities (a proxy of their capability sets). Third, we compare the relative advantage of the different groups in terms of their corresponding outcome distributions from the veil of ignorance perspective.

The reference model provides an overall (cardinal) measure of relative opportunities that obtains as a function of the dominance probabilities. There are two relevant features of this measure worth mentioning. On the one hand, the model is flexible enough to deal with categorical data. On the other hand, it can be consistently applied to any (finite) number of groups.

We have applied this methodology to the analysis of European countries with respect to equality of opportunity in education, a classic field in which this approach has shown to be fruitful. Needless to say, this methodology can be used for the analysis of many other dimensions. In the application we find that the difference in the relative education opportunities of children coming from differently educated parents is highest in Southern countries like Greece, Italy and Spain, and lowest in Northern countries like Denmark and Sweden. Spain is the only country in which the relative disadvantage in education opportunities of the children coming from the less educated fathers significantly improved when moving from the older to the younger cohorts of parents. We also obtain evidence of convergence across European countries in the relative opportunities of children coming from less and intermediately educated parents.

References

Abu-Naga, RH & Yalcin, T, 2008, Inequality measurement for ordered response health data, **Journal of Health economics**, 27(6):1614-25.

Allison, RA & Foster, J., 2004, Measuring health inequalities using qualita- tive data, **Journal of Health economics**, 23: 505-524.

Almas, I., A.W. Cappelen, J.T. Lind, E.O. Sorensen and B. Tungodden, (2011), Measuring unfair (in)equality, **Journal of Public Economics**, 95: 488- 499.

Arneson, R.J., 1989, Equality and equality of opportunity for welfare, **Philosophical Studies**, 56: 77-93.

Berman, A. & Plemmons, R.J. (1994), **Nonnegative Matrices in the Mathematical Sciences**, SIAM Classics in Applied Mathematics.

Bourguignon, F., F.H.G. Ferreira and M. Menéndez, 2007, Inequality of opportunity in Brazil, **Review of Income and Wealth**, 53: 585-618.

Checchi, D. and V. Peragine , 2010, Inequality of opportunity in Italy, **Journal of Economic Inequality**, 8: 429-450.

Chakravarty, SR & J. Silber, 2007, A generalized index of employment segregation, **Mathematical Social Sciences**, 53(2): 185-195.

Cohen, G.A., 1989, On the currency of egalitarian justice, **Ethics**, 99(4): 906-944.

Dutta, I. & Foster, J., 2011, Inequality of happiness in the US: 1972-2008, The University of Manchester.

Echenique, F., & Fryer, RG., 2005, On the measurement of segregation, Labor and Demography, Econ WPA

Fleurbaey, M., 2008, **Fairness, responsibility and welfare**, Oxford U. Press, Oxford.

Frankel, DM. & Volij, O. (2011), Measuring school segregation, **Journal of Economic Theory,** 146(1): 1-38.

Herrero, C. (1996), Capabilities and Utilities, **Economic Design**, 2 : 69-88.

Herrero, C. & JL Pinto, 2008, Capabilities and opportunities in health, working paper, FBBVA 10.

Herrero, C. & Villar, A. (2013), On the comparison of group performance with categorical data, **PLoS ONE**, 8(12): e84784. doi:10.1371/journal.pone.0084784.

Herrero, C., Méndez, I. & Villar, A. (2014) Analysis of Group Performance with Categorical Data when Agents are Heterogeneous: The Evaluation of Scholastic Performance in the OECD through PISA, **Economics of Education Review**, 40 : 140-151.

Herrero, C. & Villar, A. (2014a), Ranking Distributions of Monotone Attributes, European University Institute working paper ECO2014-06.

Herrero, C. & Villar, A. (2014b), Equality of Opportunity with Categorical Data, **Lecturas de Economía**, 81 : 9-30.

Lefranc, A., N. Pistolesi and A. Trannoy, 2008, Inequality of opportunities vs. inequality of outcomes: Are western societies all alike?, **Review of Income and Wealth**, 54: 513-546.

Lefranc, A., N. Pistolesi 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: 1189-1207.

Lieberson, S. 1976, Rank-Sum comparisons between Groups, **Sociological Methodology**, 7: 276-291.

Moreno-Ternero, J. (2007), On the design of equal-opportunity policies, **Investigaciones Económicas**, 31 : 351-374.

OECD (2013), **PISA 2012 Results: Excellence through equity. Giving every student the chance to succeed**. http://dx.doi.org/10.1787/9789264201132-en.

Palacios-Huerta, I. & O. Volij, 2004, The Measurement of Intellectual Influence, **Econometrica**, 72 (3): 963-977.

Peragine, V. and L. Serlenga 2008, Higher education and equality of oppor- tunity in Italy, **Research on Economic Inequality**, 16: 67-97.

Roemer, J.E., 1993, A pragmatic theory of responsibility for the egalitarian planner, **Philosophy and Public Affairs**, 22(2): 146-166.

Roemer, J.E., 1998, Equality of opportunity, Harvard U. Press, New York.

Sen, A., 1985, **Commodities and capabilities**, Oxford University press, Amsterdam: North Holland.

Slutzki, G., & Volij, O. (2006), Scoring of web pages and tournaments, **Social Choice and Welfare,** 26 (1): 75-92

Thomson, W. Individual and Collective Opportunities, **International Journal of Game Theory,** 16: 245-252, 1987

Thomson, W. (1994), Notions of equal or equivalent opportunities, **Social Choice and Welfare**, 11 : 137-156.

Thomson, W. (2001), On the axiomatic method and its recent applications to game theory and resource allocations, **Social Choice and Welfare**, 18 : 327-386.

Thomson, W. (2011), Consistent allocation rules", mimeo.

Thomson, W. (2012), On the axiomatics of resource allocation: interpreting the consistency principle, **Economics and Philosophy***,* 28 : 385-421

Van de gaer, D., 1993, Equality of Opportunity and Investment in Human Capital, Ph.D. dissertation. KULeuven, Leuven.

Villar, A. (2014), Education and Cognitive Skills in the Spanish Adult Population. Intergenerational Comparison of Mathematical Knowledge form PIAAC Data, **Advances in Social Sciences Research Journal**, 1 : 72-88.

Yalonetzky, G. (2012), A Dissimilarity Index of Multidimensional Inequality of Opportunity, **The Journal of Economic Inequality**, 10: 343-373

Acknowledgements: Thanks are due to Dirk Van de gaer, X. Ramos, F. Goerlich and the participants in different seminars for helpful comments. Financial support from the Spanish Ministerio de Economía y Competitividad, under projects ECO2011-29751 and ECO2013-43526-R are gratefully acknowledged.