serie ad

WP-AD 2012-17

On the comparison of group performance with categorical data

Carmen Herrero and Antonio Villar

Ivie

Working papers

Working papers

Working papers

Versión: agosto 2012 / Version: August 2012

Edita / Published by:
Instituto Valenciano de Investigaciones Económicas, S.A.
C/ Guardia Civil, 22 esc. 2 1º - 46020 Valencia (Spain)

WP-AD 2012-17

On the comparison of group performance with categorical data[*]

Carmen Herrero and Antonio Villar[**]

Abstract

This paper provides a criterion to evaluate the overall performance of a given number of groups whose members can be of different types, out of the analysis of their relative frequencies. Assuming that types can be ordered from best to worse, the starting point is that of dominance relations in pair-wise comparisons. We say that group *i* dominates group *j* when the expected type of a member of *i* is higher than the expected type of a member of *j*. To extend this principle to multi-group comparisons we introduce the notion of relative advantage of a group and show that there is a unique evaluation function that ranks consistently all groups in terms of this criterion. This function associates, to each evaluation problem, the (unique) dominant eigenvector of a matrix whose entries describe the dominance relations between groups in pairwise comparisons. An application to education illustrates the working of this model.

JEL classification numbers: D01, D70, Z13.

Keywords: Group comparisons, categorical data, dominance relationships, relative advantage, worth of a group.

[*] Acknowledgements: Thanks are due to Miguel Ángel Ballester and the participants in different seminars for helpful comments. This paper was written while the first author was visiting the IAE, under the Salvador de Madariaga project PR2010-0124 and finished while the second author was visiting the Department of Mathematics for Decisions at the University of Florence. Our research is covered by the projects SEJ2007-62656, SEJ-6882/ECON and ECO2010-21706 from the Spanish Ministry of Science and Technology, the Junta de Andalucía and the FEDER funds.

[**] C. Herrero: Universidad de Alicante, Ivie and IAE (corresponding author: carmen.herrero@ivie.es). A. Villar: Pablo de Olavide University and Ivie.

1. Introduction

There are many situations in which one is interested in making group comparisons in terms of some ordered characteristic. Think for instance of the evaluation of different countries out of the data of *Self Assessed Health status* surveys or out of data concerning the educational levels achieved. In the first case people usually declare that their health situation is one among four or five categories, ranging from “very bad” to “excellent”. In the second case we find a distribution of the population of each country into different categories, from no education to university studies. Comparing societies out of those data requires either finding a way of attaching numbers to the different categories (health statuses or education levels in our examples) and summarizing those values by some aggregation procedure, or else devising a criterion that is capable of dealing with qualitative data.

We shall follow here the latter approach. More precisely, this paper presents a procedure to make group comparisons when the traits or achievements of their members are described by ordered categorical data (e.g. age intervals, income brackets, health statuses, education levels, prestige positions). The setting consists, therefore, of a finite set of groups whose members are classified into a given number of categories, called here *types*, which describe their characteristics or their realizations. Types are assumed to be ordered, so that one can unambiguously say that a type “precedes”, is “higher than”, or is “preferred to” another. Our goal is to find a suitable summary measure that synthesizes the key features of the different groups taking into account the distribution of its members along those categories. The evaluation will focus on the relative frequencies of the agents in the different cells that arise from the double partition into groups and types. Comparing groups amounts, therefore, to comparing the distribution patterns of their types.

These ideas are related to the statistical analysis dealing with the similarity between rank distributions and the sociological and economic literature dealing with segregation (see for instance Reardon & Firebaugh (2002), Grannis (2002), Reardon & O’Sullivan (2004), del Río & Alonso (2010), or Yaloneztky (2010)). An early contribution worth stressing is that of Lieberson (1976). He introduces the notion of *Net Difference* for the evaluation of pairs of groups, as the difference between the probability that an agent from group *i* be in a better position than an agent from group *j* and the probability that an agent from group *j* be in a better position than an agent from group *i*.

Here we extend this idea to a more general setting (i.e. comparing any finite number of groups). The evaluation is obtained in three steps. First, we define the *relative advantage* of group *i* with respect to group *j*, as the ratio between the probability that *i* dominates *j* and the probability that *i* be dominated by other groups. Second, we obtain the overall advantage of a group as a weighted average of its relative advantages with respect to all other groups. Finally, we select an invariant weighting system so that the weights used for the weighted average correspond, precisely, to those yielded by the overall evaluation.

The resulting evaluation function corresponds to the eigenvector of a suitably constructed matrix that incorporates the information about the distribution of the groups across the different types. The solution is thus a fixed point of a mapping that can be interpreted as the limit process of an evaluation procedure in which the weights with which we calculate the overall advantage of a group are progressively adjusted according to the outcome of the evaluation process. This type of evaluation has a similar flavour to some of the ways of evaluating the impact of scientific journals (see Pinski & Narin (1976), Laband & Piette (1994), Palacios Huerta & Volij (2004), Serrano (2004), Waltman & Jan van Eck (2010) or the construction of the *Eigenfactor*).

Our main contribution in this paper can be regarded as that of framing the evaluation problem so that we can rely on a conventional solution (the fixed point of a linear mapping) to provide the type of evaluation we are looking for. Moreover, as this fixed point corresponds to the dominant eigenvector of a Perron matrix, the proposed solution exhibits simple, useful, and well-known properties (e.g. existence, uniqueness, positiveness, stability, and regular behaviour regarding changes in the parameters). Note that using the distribution of the groups across types permits one to extract cardinal information out of categorical data.

The paper is organized as follows. Section 2 presents the model and proves the existence of the evaluation function with the required properties. Section 3 discusses some salient features of this solution and includes an elementary characterization. An application regarding the evaluation of the OECD countries out of the PISA 2009 report on educational achievement is provided in Section 4. A few final comments are gathered in Section 5.

2. The Model

Consider a set of G groups, G = {1, 2, …, g}, with $g \geq 2$, and let n_i denote the number of members in group i = 1, 2, …, g. We assume that the individual characteristics of the groups' members induce a partition in terms of *s types* or categorical positions, $C = \{c_1, c_2, \dots, c_s\}$, ordered from best to worst, $c_1 \succ c_2 \succ \dots \succ c_s$.

Let a_{ir} denote the share of members of group i with type r, i.e. $a_{ir} = n_{ir}/ n_i$, where n_{ir} is the number of members in group i with type r. Our goal is comparing the relative performance of the different groups, out of the distribution of their types. Let p_{ij} denote the *probability that a member of group i be in a better position than a member of group j.* Since types are ordered, this probability can be easily computed through the following formula:

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

Similarly, p_{ji} denotes the probability that a representative member of group j be better off than a representative member of group i. And, consequently, $e_{ij} = 1 - p_{ij} - p_{ji}$ stands for the probability that a member from i, picked at random, be at the same position than a member from j (with $e_{ij} = a_{ij}a_{j1} + \dots + a_{is}a_{js}$).

Remark 1.- *We shall assume, for the sake of ease, that all p_{ij} are strictly positive. That amounts to excluding the existence of a group all whose members are of the lowest type.*

Consider now the following:

- Definition 1: We say that group i ***dominates*** group j in a pairwise comparison whenever it is more likely that picking at random a member from i she will be in a higher position than a member from j randomly chosen. That is, $i \succ j \Leftrightarrow p_{ij} > p_{ji}$.

When there are only two groups involved this is a sound criterion that permits one to evaluate their relative performance in an unambiguous way. This type of pairwise comparison is reminiscent of Lieberson's (1976) *Index of Net Difference*.[1] Extending this

[1] The Index of Net Difference, $ND(i, j)$, is defined as the absolute value of the difference $p_{ij} - p_{ji}$. This index provides an estimate of the relative advantage of any two groups. When $ND(i, j) = 0$ a member chosen at random from group i has the same probability of being better off than a member picked at random from group j than the other way around. On the opposite extreme, we find the case $ND(i, j) = 1$, which happens whenever all members of one group occupy better positions than those in the other. Intermediate cases generate values in the interior of the interval [0, 1].

principle to a more general setting, involving any finite number of groups, is non-trivial and requires some additional elaboration. We have to devise a way of comparing the relative position of members in each group with respect to all other groups.

Let P denote the set of all pairwise comparisons p_{ij}, $i \neq j$. This set fully describes the data of our evaluation problem. Let us introduce the notion of *relative advantage* of a group with respect to another, as follows:

- Definition 2: Given a problem P, the ***relative advantage*** of group i with respect to group j, $\pi_{ij}(P)$, is given by:

$$\pi_{ij}(P) = \frac{p_{ij}}{\sum_{k \neq i} p_{ki}} \qquad [2]$$

That is, π_{ij} is the probability that group i dominates group j, in a pairwise comparison, relative to the aggregate probability of group i being dominated by some other group. When there are only two groups involved, we have $\pi_{ij} = p_{ij}/p_{ji}$, so that $\pi_{ij} > 1$ indicates that i dominates j. Moreover, we find that, for all pairwise comparisons with any given number of groups,

$$\frac{\pi_{ij}(P)}{\pi_{ik}(P)} = \frac{p_{ij}}{p_{ik}}$$

that is, the ratio of the relative advantage of i with respect to j and with respect to k coincides with the ratio of their associated domination probabilities. This is not the case, however, in general as one should expect that:

$$\frac{\pi_{ij}(P)}{\pi_{kj}(P)} \neq \frac{p_{ij}}{p_{kj}}$$

due to the effect of the different degree in which groups i and k are dominated by other groups.

If π_{ij} is the relative advantage of group i with respect to group j, what can we say about the overall performance of group i? The simplest way of getting such a global evaluation is by assigning to each group a weighted average of its relative advantages. That is,

$$\mu_i = \sum_{j \neq i} \lambda_j \pi_{ij}(P), \qquad \sum_{j=1}^{g} \lambda_j = 1 \qquad [3]$$

where $\lambda_j > 0$ is a measure of the importance attached to j (*input relevance*) and μ_i is the resulting overall evaluation of group i (*output relevance*). Equation [3] may thus be regarded as transforming the relevance initially assigned to the different groups, $(\lambda_1, \lambda_2, ..., \lambda_g)$, into new evaluations, $(\mu_i, \mu_2, ..., \mu_g)$, by taking into account their relative advantages. The natural question is whether we can find an *invariant* system of weights. That is, a way of attaching the relevance of the different groups, $\left(\lambda_1^*, \lambda_2^*, ..., \lambda_g^*\right)$, so that:

$$\left(\mu_1, \mu_2, ..., \mu_g\right) = \left(\lambda_1^*, \lambda_2^*, ..., \lambda_g^*\right)$$

This property would ensure a consistent evaluation, in the sense that the importance attached to the different groups derives, precisely, from the importance that the evaluation function yields.

From a formal point of view the existence of such a special weighting system corresponds to a fixed point of the mapping H that transforms input relevance into output relevance, $(\mu_i, \mu_2, ..., \mu_g) = H(\lambda_1, \lambda_2, ..., \lambda_g)$, where $H_i(\lambda_1, \lambda_2, ..., \lambda_g) = \sum_{j \neq i} \lambda_j \pi_{ij}(P)$, i = 1, 2, …, g. Such a fixed point, call it $\mathbf{v} = H(\mathbf{v})$, associates to each group a weighted sum of its relative advantages, where the weights correspond to the importance that function H determines. That is, $v_i = \sum_{j \neq i} \pi_{ij}(P) v_j$. We call the value v_i so obtained the **worth** of group i and define the following:

- Definition 3: A ***consistent evaluation function*** is a mapping F that associates, to each group i = 1, 2, …, g, in an evaluation problem P, its worth. That is, for each problem P we have: $F(P) = \mathbf{v}$ with:

$$v_i = \sum_{j \neq i} \pi_{ij}(P) v_j, \quad i = 1, 2, ..., g \qquad [4]$$

The worth of group i, relative to an evaluation function F, is the weighted average of the relative advantage of this group with respect to all other groups, where the weights correspond to their associated valuations. From this it follows that the worth of a group is higher, other things equal, the higher the value of the groups it dominates.

The case in which there are only two groups involved has an interesting property: the ratio between their valuations coincides with the ratio of the probability of one dominating the other. That is,

$$\frac{v_i}{v_j}=\frac{\pi_{ij}(P)v_j}{\pi_{ji}(P)v_i}\Rightarrow\frac{\pi_{ij}(P)}{\pi_{ji}(P)}=\left(\frac{v_i}{v_j}\right)^2=\frac{p_{ij}/p_{ji}}{p_{ji}/p_{ij}}\Rightarrow\frac{v_i}{v_j}=\frac{p_{ij}}{p_{ji}}$$

As pointed out before, finding a consistent evaluation vector amounts to finding a fixed point of function *H*, for each possible problem. We show next that such a fixed point always exists and it is unique (once the scale has been chosen and bearing in mind the positiveness assumption in Remark 1). [2]

Theorem 1.- *Let P be an evaluation problem regarding* $g \geq 2$ *groups whose members are classified into s ordered types. There exists a unique consistent evaluation function F, with F(P) =* $\mathbf{v} >> 0$, *with* $v_i=\sum_{j\neq i}\pi_{ij}(P)v_j,\ \ i=1, 2, ..., g$, *and* $\sum_{i=1}^{g} v_i=1$.

<u>Proof.-</u>

Consider now that the information relative to problem *P* is arranged in the form of a $g\times g$ matrix *P** whose (*i, j*) entry is p_{ij}, for $i \neq j$. The diagonal elements of that matrix, called D_i, are given by $D_i=\sum_{j\neq i}\left(p_{ij}+e_{ij}\right)$, that is, the probability of a member of group *i* be better off than or equal to a member picked at random on the other groups. We thus have:[3]

$$P^{*}=\begin{bmatrix} D_1 & p_{12} & \cdots & p_{1g} \\ p_{21} & D_2 & \cdots & p_{2g} \\ \cdots & \cdots & \cdots & \cdots \\ p_{g1} & p_{g2} & \cdots & D_g \end{bmatrix} \qquad [5]$$

Matrix *P** is simply a particular way of arranging the information concerning the problem under consideration. Now observe that *P** is a square matrix with positive entries (i.e. a Perron matrix). Moreover, by construction, all columns of *P** add up to (g – 1). To see that notice that $D_i=\sum_{j\neq i}\left(p_{ij}+e_{ij}\right)=\left(g-1\right)-\sum_{j\neq i}p_{ji}$. Therefore, *P** has a single

[2] The uniqueness and strict positivity of the weighting system only requires the irreducibility of matrix P* defined below. The strict positivity of all p_{ij} is a sufficient condition for that.

[3] To understand better what it says, take for instance the first column. The entries (2,1), (3,1), …, (g,1) tell us the probability of a member of group 1 to be worse-off than a member of group 2, the probability of a member of group 1 to be worse-off than a member of group 3, and so on. The term D_1 refers to the complementary state, that is, the probability that group 1 weakly dominates other groups.

dominant positive eigenvalue, equal to (g – 1), that has associated a strictly positive eigenvector $\mathbf{v} >> \mathbf{0}$, with:

$$P^*\mathbf{v} = (g-1)\mathbf{v} \qquad [6]$$

This eigenvector, $\mathbf{v} = (v_1, v_2, \ldots, v_g)$, is unique up to a scalar multiplication, so that we can assume, without loss of generality, that $v_1 + v_2 + \ldots + v_g = 1$. Observe that the jth entry of that eigenvector can be written as:

$$v_i = \sum_{j \neq i} \frac{p_{ij} v_j}{\sum_{j \neq i} p_{ji}} \qquad [7]$$

The evaluation function is thus implicitly defined as follows:

$$F(P) = \left\{ \mathbf{v} \in \Re^g_{++} \;/\; P^*\mathbf{v} = (g-1)\mathbf{v},\; \sum_{i=1}^{g} v_i = 1 \right\}$$

Q.e.d.

The evaluation function F tells us the worth of each group for a given problem. The worth of a group refers to the situation of a representative member vis a vis the representative members of all other groups (it is, therefore, a vector of values *independent on the groups sizes*). Function F allows comparing consistently the relative advantage of any two groups: $v_i > v_j$ means that members of i are on average in a better position than members of j.

Note that the evaluation formula in Theorem 1 satisfies the property of *Stochastic Dominance*: when a group exhibits better values than another in all categories it gets a higher worth. That is, if the distribution of group i stochastically dominates the distribution of j, then $v_i > v_j$. A particular consequence of this property is that the worth of a group all whose members are of the lowest type will be equal to zero.

3. The worth vector from a different angle

The worth vector $\mathbf{v}$ in Theorem 1 satisfies a number of interesting properties that reinforce the operational and normative appeal of this evaluation procedure. Here we provide two different views of this evaluation formula that stress its relevance and add new insights on the nature of the approach. First, we present the worth vector as a way of extending the dominance criterion form the two-group case to the general setting, without missing the indirect relations that are lost in simple pair-wise comparisons. Second, we look at the worth vector as the limit of an evaluation process in which the evaluation of the groups are adjusted sequentially.

3.1. From two to many

One alternative way of looking at the worth vector is as a linear extension of the principle according to which in the two-group case the relative worth of both groups coincides with the ratio of their relative advantages (that is, $v_1/v_2 = p_{12}/p_{21}$). Let us call this property *Proportionality in the two-group case*.

When there are many groups, g > 2, pairwise comparisons are not informative enough as they miss the indirect dominance relations. Yet one can apply the Proportionality in the two-group case principle in this context as follows. Given a problem P involving g > 2 groups, construct g problems of dimension 2 by confronting each of the initial groups with an artificial group consisting of "all other groups". That is, we construct problem P_i , i=*1,…g,* where problem P_i consists of group i and the aggregate of the remaining (g-1) groups. Let us call (ω_i, ω_{-i}) the evaluation vector associated to problem P_i, where ω_i represents the valuation of group i in this problem, and ω_{-i} stands for the valuation of the complementary group. The Proportionality in the two-group case implies here that, for all $i=1, \ldots, g,$

$$\frac{\omega_i}{\omega_{-i}} = \frac{\sum_{j \neq i} p_{ij}}{\sum_{j \neq i} p_{ji}}$$

We say that an evaluation function satisfies the property of ***Agreement*** when, for any problem P involving g groups, for all pair-wise decompositions P_i, all $i = 1, \ldots, g,$ we have:

$$\frac{\omega_i}{\omega_{-i}} = \frac{\sum_{j \neq i} p_{ij}}{\sum_{j \neq i} p_{ji}}$$

Clearly, agreement implies Proportionality in the two-group case.

It is easy to see that our evaluation function in Theorem 1 satisfies this property. Moreover, this property together with the standard properties of Anonymity (permuting the ordering of the groups simply yields a corresponding permutation of the components of vector **v**) and *Replication Invariance* (replicating a group does not change its worth, as only the shares enter the evaluation function), characterizes such an evaluation function.

Formally:

Theorem 2: *Let P be an evaluation problem regarding* $g \geq 2$ *groups whose members are classified into s ordered types. An evaluation function satisfies anonymity, replication invariance, and agreement if and only if it is the consistent evaluation function F, with F(P) =* $\mathbf{v} >> 0$*, with* $v_i = \sum_{j \neq i} \pi_{ij}(P) v_j$, $i = 1, 2, ..., g$, *and* $\sum_{i=1}^{g} v_i = 1$.

Proof:

It is enough to see the "only if" part. Suppose that an evaluation function satisfies those properties. Replication invariance indicates that *F* only depends on the relative proportions. Anonymity says that we can only rely on the data provided by matrix *P**. Agreement relates the value attached to any group to the evaluation of the *g* bi-dimensional reduced problems. Then, for all *i = 1, …, g,*

$$\frac{\omega_i}{\omega_{-i}} = \frac{\sum_{j \neq i} p_{ij}}{\sum_{j \neq i} p_{ji}}$$

Now observe that we can write:

$$\frac{\omega_i}{\omega_{-i}} = \frac{\sum_{j \neq i} p_{ij}}{\sum_{j \neq i} p_{ji}} = \frac{v_i \sum_{j \neq i} p_{ij}}{\sum_{j \neq i} p_{ij}\, v_j}$$

and from here we get that:

$$v_i = \frac{\sum_{j \neq i} p_{ij} v_j}{\sum_{j \neq i} p_{ji}}.$$

Q.E.D.

3.2. The worth vector as the limit of an adjustment process

The fact that the worth vector is the eigenvector associated to the dominant eigenvalue of matrix *P** permits one to view this evaluation formula from still another perspective. Namely, the worth vector is the limit of a dynamic evaluation process in which the worth of a group is adjusted sequentially by using matrix *P**, starting from an arbitrary evaluation of the different groups.

That is, let us take an initial evaluation vector, ω, and proceed as follows:[4]

$$\omega^{(1)} = P * \omega$$
$$\omega^{(2)} = P * \omega^{(1)}$$
$$\ldots$$
$$\omega^{(t)} = P * \omega^{(t-1)}$$
$$\ldots$$

Then we end up by getting:

$$\mathbf{v} = \lim_{t \to \infty} \omega^{(t)}$$

From this viewpoint the evaluation function appears as the final step of an evaluation process that adjusts progressively the importance of the different groups out of the result of the former evaluation.

4. An application: The evaluation of compulsory education in the OECD through PISA 2009

The Programme for International Student Assessment (PISA) provides the broadest dataset for the evaluation of schoolchildren performance and the characteristics of their schooling and family environments. It is a triennial worldwide test of 15-year-old schoolchildren's scholastic performance, the implementation of which is coordinated by the OECD. PISA surveys started in 2000 with the aim of evaluating the students' ability, about the end of compulsory education, in three different domains: reading, mathematics and science. Every period of assessment specialises in one particular category, but it also tests the other two main areas studied. The 2009 report, in which almost half a million students completed the assessment in the 65 participating countries and large economies, has focused on reading abilities (as it was the case in 2000).

Reading literacy is a key aspect of individuals' learning ability at school and conditions their participation in social life. "Levels of *reading literacy* are more reliable predictors of economic and social wellbeing than is the quantity of education as measured by years at school or in post-school education… It is the quality of learning outcomes, not the length of schooling, that makes the difference." (PISA report vol. I, p. 32).

[4] One may think, in order to help the intuition, of the special case in which $\omega_i = 1/g$ so that the initial evaluation of a group corresponds to the arithmetic mean of its relative advantages.

One of the assets of the PISA report is that it provides a unified scoring system to evaluate the performance of 15-year-old students in very different countries. The units of those scores are set with respect to the values obtained in the 2000 wave of the report, by taking a value of 500 for the average of the OECD Member States with a standard deviation of 100. Besides the average score the PISA report classifies the students in six categories defined by a gradual increase of reading competence. Each of those levels is defined in terms of the capacity of the students to achieve certain cognitive processes and operationalized in term of ranges of the scores obtained by the students (see Figure 1.2.12 in volume I of the PISA report for details). Table 1 summarizes the scoring intervals that parameterize those levels of competence.

Table 1. Levels of reading competence

Level of competence		Score range	% of OECD students within the level
Level 6		> 698	0.8
Level 5		626 - 698	6.8
Level 4		553 - 626	20.7
Level 3		480 - 553	28.9
Level 2		407 - 480	24.0
Level 1	Level 1a	335 - 407	13.1
	Level 1b	262 - 335	4.6

The distribution of the mean scores among the countries shows a relatively low dispersion, with a coefficient of variation of 0.1102 for the 65 participating economies (a figure that reduces to 0.046 for the OECD countries).[5] Differences seem to be much more important when one looks at the distribution of the students within the different levels of competence, as reported in Table 2 for the OECD countries.

[5] Yet, the difference between top and bottom performers is huge: there are 242 score points of difference between Shangai-China and Kyrgyzstan, which corresponds to six formal years of schooling. The difference between the top and the bottom OECD countries (Korea and Mexico, respectively) is of 114 score points, more than the equivalent of two school years.

Table 2. Share of students at each proficiency level on the reading scale (OECD)

	Level 1 (or below)	Level 2	Level 3	Level 4	Level 5	Level 6
Australia	0.142	0.204	0.285	0.241	0.107	0.021
Austria	0.276	0.241	0.260	0.174	0.045	0.004
Belgium	0.177	0.203	0.258	0.249	0.101	0.011
Canada	0.103	0.202	0.300	0.268	0.110	0.018
Chile	0.306	0.332	0.256	0.093	0.013	0.000
Czech Republic	0.231	0.274	0.270	0.174	0.047	0.004
Denmark	0.152	0.260	0.331	0.209	0.044	0.003
Estonia	0.133	0.256	0.338	0.212	0.054	0.006
Finland	0.081	0.167	0.301	0.306	0.129	0.016
France	0.198	0.211	0.272	0.224	0.085	0.011
Germany	0.185	0.222	0.288	0.228	0.070	0.006
Greece	0.213	0.256	0.293	0.182	0.050	0.006
Hungary	0.176	0.238	0.310	0.216	0.058	0.003
Iceland	0.168	0.222	0.306	0.219	0.075	0.010
Ireland	0.172	0.233	0.306	0.219	0.063	0.007
Israel	0.265	0.225	0.255	0.181	0.064	0.010
Italy	0.210	0.240	0.289	0.202	0.054	0.004
Japan	0.136	0.180	0.280	0.270	0.115	0.019
Korea	0.058	0.154	0.330	0.329	0.119	0.010
Luxembourg	0.260	0.240	0.270	0.173	0.052	0.005
Mexico	0.401	0.330	0.212	0.053	0.004	0.000
Netherlands	0.143	0.247	0.276	0.235	0.091	0.007
New Zealand	0.143	0.193	0.258	0.248	0.129	0.029
Norway	0.150	0.236	0.309	0.221	0.076	0.008
Poland	0.150	0.245	0.310	0.223	0.065	0.007
Portugal	0.176	0.264	0.316	0.196	0.046	0.002
Slovak Republic	0.222	0.281	0.285	0.167	0.042	0.003
Slovenia	0.212	0.256	0.292	0.193	0.043	0.003
Spain	0.196	0.268	0.326	0.177	0.032	0.002
Sweden	0.174	0.235	0.298	0.203	0.077	0.013
Switzerland	0.168	0.227	0.297	0.226	0.074	0.007
Turkey	0.245	0.322	0.291	0.124	0.018	0.000
United Kingdom	0.184	0.249	0.288	0.198	0.070	0.010
United States	0.176	0.244	0.276	0.206	0.084	0.015
OECD total	0.198	0.244	0.279	0.199	0.070	0.010
OECD average	0.188	0.240	0.289	0.207	0.068	0.008

This is an evaluation problem in which our model can help. Here groups are OECD countries, members are the 15-year old students within each country, and categories correspond to levels of reading competence. Table 3 below gives the calculation of the worth of the different countries and compares those values with the mean scores of the PISA tests. The difference between those evaluation procedures ranges from + 83 % for

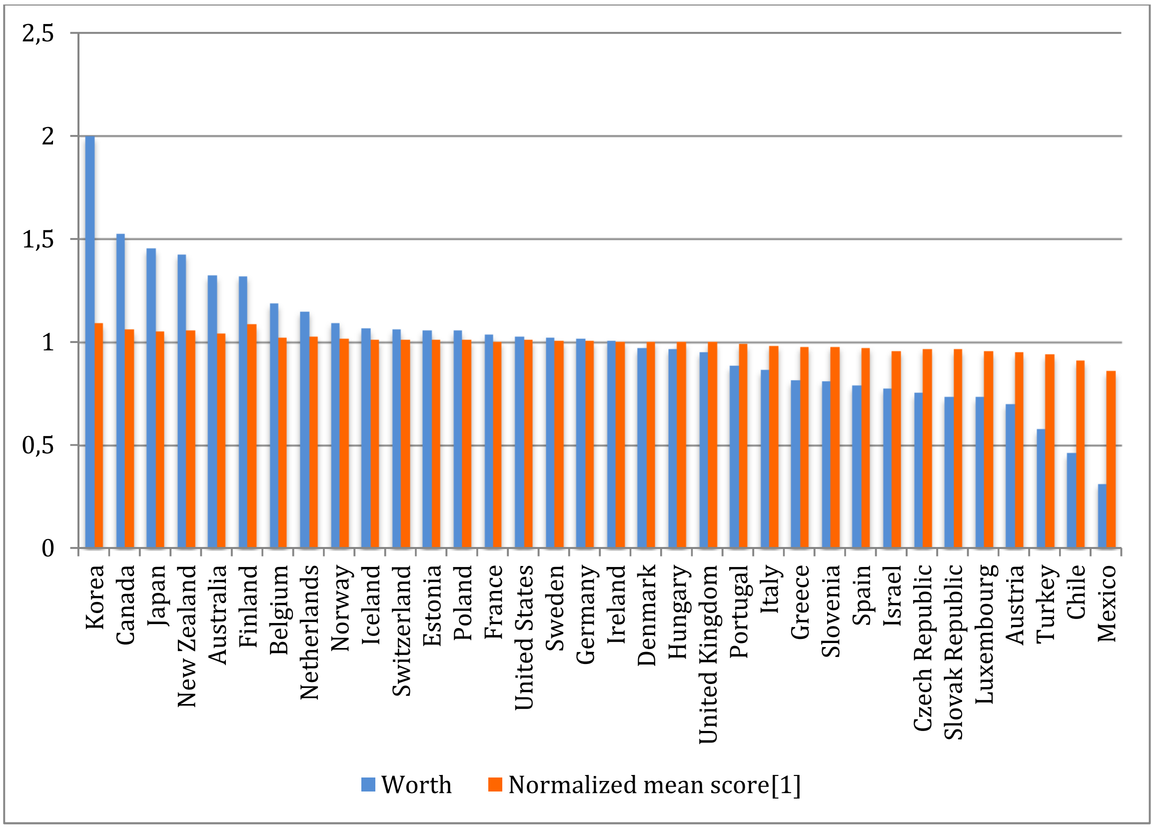
Korea to – 63 % for Mexico. The coefficient of variation of the worth values is 0.32, almost seven times that of the mean scores of the tests. The pictures we get concerning the relative positions of the countries is, therefore, rather different.

Table 3. Worth and mean scores of the OECD countries (reading competence, PISA 2009)

Country	**(A) Worth**	**(B) Normalized mean score**[6]	**(C) % Difference 100x(A-B)/B**
Australia	1.326	1.045	26.89
Austria	0.704	0.953	-26.13
Belgium	1.190	1.026	15.98
Canada	1.525	1.063	43.46
Chile	0.464	0.911	-49.07
Czech Republic	0.755	0.970	-22.16
Denmark	0.972	1.004	-3.19
Estonia	1.058	1.016	4.13
Finland	1.322	1.087	21.62
France	1.038	1.006	3.18
Germany	1.019	1.008	1.09
Greece	0.817	0.980	-16.63
Hungary	0.971	1.002	-3.09
Iceland	1.069	1.014	5.42
Ireland	1.009	1.006	0.30
Israel	0.780	0.961	-18.83
Italy	0.867	0.986	-12.07
Japan	1.456	1.055	38.01
Korea	1.999	1.093	82.89
Luxembourg	0.735	0.957	-23.20
Mexico	0.315	0.862	-63.46
Netherlands	1.152	1.030	11.84
New Zealand	1.428	1.057	35.10
Norway	1.096	1.020	7.45
Poland	1.058	1.014	4.34
Portugal	0.886	0.992	-10.69
Slovak Republic	0.737	0.968	-23.86
Slovenia	0.813	0.980	-17.04
Spain	0.791	0.976	-18.95
Sweden	1.025	1.008	1.69
Switzerland	1.063	1.016	4.63
Turkey	0.579	0.941	-38.47
United Kingdom	0.954	1.002	-4.79
United States	1.028	1.014	1.38

[6] The OECD mean score is set equal to 1.

Figure 1. Worth and mean scores values of the OECD countries in reading competence in PISA 2009 (OECD = 1)



5. Final Comments

There are many different evaluation problems that involve several groups whose members are classified into ordered types. The solution proposed here exploits the information on the distribution of their members across types in order to provide a quantitative estimate of their relative situation. We frame the problem in such a way that the solution corresponds to the eigenvector of a suitable Perron matrix. The underlying evaluation principle is that of comparing representative members of the different groups. When the expected type of a member of group *i* is higher than the expected type of a member of group *j*, we consider that group *i* is in a better position than group *j*. The way of constructing such an evaluation function is, furthermore, justified on the basis of the properties such an evaluation fulfils, and it is axiomatically characterized.

We have provided an illustration regarding the evaluation of the educational performance of the 15-year old students in the OECD countries, using the data of PISA 2009. This allowed us to obtain a deeper insight on this performance, that in some aspect differ from that obtained by the traditional methods.

Our instrument is flexible enough to be applied to different settings, as for example, to the evaluation of health achievements, the analysis of gender discrimination, or the relative performance of the job market in different countries. An application to the analysis of equality of opportunity can be found in Herrero, Méndez & Villar (2012).

References

Eigenfactor, www.eigenfactor.org

Grannis, R. (2002), Segregation Indices and their Functional Inputs, **Sociological methodology**, 32 (1), 69-84

Herrero, C., Méndez, I. & Villar, A. (2012), Relative Equality of Opportunity: A Statistical Approach, INDICSER working paper.

Laband, D.N., & Piette, M.J. (1994). The relative impacts of economics journals: 1970–1990. **Journal of Economic Literature**, 32(2), 640–666.

Lieberson, S. (1976) Rank-sum comparisons between groups, **Sociological Methodology** 7, 276-291

OECD (2010a), **PISA 2009 Results: What Students Know and Can Do** (Volume I). http://dx.doi.org/10.1787/9789264091450-en

OECD (2010b), **PISA 2009 Results: Overcoming Social Background** (Volume II). http://dx.doi.org/10.1787/9789264091504-en

Reardon, S. F. & Firebaugh, G. (2002), Measures of Multi-Group Segregation, **Sociological Methodology**, 32 : 33-76.

Reardon, S. F. & O'Sullivan, D. (2004), Measures of Spatial Seggregation, **mimeo**, Pennsylvania State University.

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

Pinski, G., & Narin, F. (1976). Citation influence for journal aggregates of scientific publications: Theory, with application to the literature of physics. **Information Processing and Management**, 12(5), 297–312.

Serrano, R. (2004). The measurement of intellectual influence: The views of a sceptic. Economics Bulletin, 1(3), 1–6.

del Río, C. & Alonso Villar, O. (2010), Occupational Segregation Measures: A Role for Status, **Ecineq working paper** 2010 – 167.

Villar, A. (2011), The Educational Development Index, **mimeo**.

Waltman, L. & Jan van Eck, N. (2010), The Relation Between Eigenfactor, Audience Factor, and Influence Weight, **Journal of the American Society for Information Science and Technology**, 61 : 1476-1486.

Yalonetzky, G. (2010), A Dissimilarity Index of Multidimensional Inequality of Opportunity, **OPHI working paper** nº 39