serie ad

WP-AD 2020-01

Dealing with categorical data in a multidimensional context: The Multidimensional Balanced Worth

Carmen Herrero and Antonio Villar

Ivie

Working papers

Working papers

Working papers

Los documentos de trabajo del Ivie ofrecen un avance de los resultados de las investigaciones económicas en curso, con objeto de generar un proceso de discusión previo a su remisión a las revistas científicas. Al publicar este documento de trabajo, el Ivie no asume responsabilidad sobre su contenido.

Ivie working papers offer in advance the results of economic research under way in order to encourage a discussion process before sending them to scientific journals for their final publication. Ivie's decision to publish this working paper does not imply any responsibility for its content.

La edición y difusión de los documentos de trabajo del Ivie es una actividad subvencionada por la Generalitat Valenciana, Conselleria de Hacienda y Modelo Económico, en el marco del convenio de colaboración para la promoción y consolidación de las actividades de investigación económica básica y aplicada del Ivie.

The editing and dissemination process of Ivie working papers is funded by the Valencian Regional Government's Ministry for Finance and the Economic Model, through the cooperation agreement signed between both institutions to promote and consolidate the Ivie's basic and applied economic research activities.

La Serie AD es continuadora de la labor iniciada por el Departamento de Fundamentos de Análisis Económico de la Universidad de Alicante en su colección "A DISCUSIÓN" y difunde trabajos de marcado contenido teórico. Esta serie es coordinada por Carmen Herrero.

The AD series, coordinated by Carmen Herrero, is a continuation of the work initiated by the Department of Economic Analysis of the Universidad de Alicante in its collection "A DISCUSIÓN", providing and distributing papers marked by their theoretical content.

Todos los documentos de trabajo están disponibles de forma gratuita en la web del Ivie http://www.ivie.es, así como las instrucciones para los autores que desean publicar en nuestras series.

Working papers can be downloaded free of charge from the Ivie website http://www.ivie.es, as well as the instructions for authors who are interested in publishing in our series.

Versión: abril 2020 / Version: April 2020

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

DOI: http://dx.medra.org/10.12842/WPAD-2020-01

WP-AD 2020-01

Dealing with categorical data in a multidimensional context: The Multidimensional Balanced Worth*

Carmen Herrero and Antonio Villar**

Abstract

This paper refers to the evaluation of the relative performance of a set of populations, in a multidimensional context, by comparing outcome distributions relative to categorical variables (qualitative or quantitative data). This type of problem appears in many different fields, such as Medicine, Social Sciences or Engineering. We address this family of problems by extending the balanced worth (Herrero & Villar, 2018) to a multidimensional setting. To illustrate the working of this evaluation protocol, we analyse the human capital of the young in nine selected European countries, from three different viewpoints.

Keywords: relative performance, categorical variables, multidimensional evaluation.

JEL classification numbers: D71.

---

* Thanks are due to Héctor García Peris (Ivie) for programming the algorithm, and IMÉRA. This work was supported by the Spanish Ministry of Economia, Industria y Competitividad (projects ECO 2015-65820 and ECO 2015-65820P), and Generalitat Valenciana (PROMETEO 2019/037).

** C. Herrero: University of Alicante, FAE and Ivie. A. Villar: Universidad Pablo de Olavide and Ivie.
Correspondence to: Carmen Herrero, email: Carmen.herrero@ua.es

1 Introduction

We present here a methodology that permits one comparing the relative performance of different populations concerning several dimensions when the data are categorical. More specifically, the type of evaluation problems we address here is that in which the variables that measure each dimension correspond to distributions of the populations into levels of performance. Those levels of performance, or categories, can be qualitative (e.g. members of the population with a given characteristic) or quantitative (intervals of the range of a reference variable).

Let us illustrate those types of problems by considering one particular example. Suppose we have to evaluate a medical trial regarding the efficacy of alternative treatments for computer workers with neck pain. Subjects are gathered into different groups with similar features. We can think of four populations, two with varying protocols of physiotherapy, one with pain relievers and muscle relaxing pills and the last one with placebo tablets serving as the control group. Those procedures are to be compared regarding three dimensions: pain reduction, disability improvement, and endurance enhancement. The first dimension is commonly measured in terms of a 0-10 pain scale. The second dimension is often a qualitative (self-perceived) variable. The last dimension can be measured quantitatively in terms of time enduring some standardized effort. Of course, we measure those variables before and after the treatment for each group.

In this example, the populations to be compared are those four groups of computer workers with neck pain, subject to different treatments. We compare the groups in three dimensions, each one dimension measured by a variable that provides the distribution of each population into a series of levels. The results regarding the first dimension, pain, are registered into eleven levels (from 0 to 10). Note that the variable that measures this dimension is intrinsically qualitative, even though expressed in numerical terms. The results for the second dimension, disability, is given in a qualitative variable (e.g. very high, high, medium, low, and very low). Finally, the third dimension, endurance, can be recorded in time intervals. It is also interesting to observe that we may be willing to evaluate differences by types of subjects (e.g. by age intervals), which also introduces the question of the heterogeneity of those populations.

This type of problem was analysed in Herrero & Villar (2013) for the case of a single dimension. They introduced the notion of *worth* to compare the achievements of different populations when their outcomes are described in terms of ordered categorical variables. The key principle for making those comparisons was that of ranking higher those populations for which it is more likely, on average, to get better outcomes. The resulting evaluation obtained as the dominant eigenvector of a matrix whose entries describe the dominance relations between populations in pair-wise comparisons. This vector is the stable distribution of the winners, in the long run, of a series of tournaments between the populations; its entries provide the evaluation of their relative performance. This work was extended in Herrero & Villar (2018) taking specifically into account the presence of ties in those pairwise comparisons and the existence of heterogeneous agents, aspects not included in the original work. That generalized version of the worth was called *balanced worth*.

The use of stable distributions to compare the performance of different populations in terms of a single variable appears in a diversity of fields. Freeman (1977), Wasserman & Faust (1994), or Newmann (2003) apply it to the study of centrality in networks. Pinsky & Narin (1976), Liebowitz & Palmer (1984), Palacios-Huerta & Volij (2004) use it to evaluate the relevance of the journals in citation analysis, and Albarrán et al. (2017) to compare the performance of countries in different disciplines based upon the number of citations. Keener (1993), or Slutzski & Volij (2005) apply it to the ranking of teams in competitions. Pifarré i Arolas & Dudel (2019) employ this approach to the measurement of population health. Stable distributions also appear in the literature regarding tournaments, that goes back to Daniels (1969) and Moon & Pullman (1970) (see also Laslier 1997, Saaty 2003, Slutsky & Volij 2006 and Boccard 2020). Not to mention the Eigenfactor, a compelling alternative to raw citation impact indices (http://www.eigenfactor.org), or the well-known Google Page-rank protocol (Brin & Page 1998), perhaps the most widespread algorithm in history.

In this paper, we address the problem of applying this evaluation procedure in a multidimensional context so that we can deal with the types of issues described in the example above. In particular, we provide in Section 2 a natural extension of the balanced worth to the evaluation of problems involving several dimensions (what we call *multidimensional balanced worth*). We can informally describe this extension as follows. Populations are randomly selected and compared within each dimension by confronting representative agents in an indefinite process that also involves a random selection of the dimension to be considered. Our work includes free access to an on-line algorithm that performs all the required computations instantly.

We also provide an empirical application of this methodology, in Section 3. It deals with the comparison of the human capital of young adults for nine selected European countries (Finland, France, Germany, Greece, Ireland, Italy, Netherlands, Spain and the United Kingdom), as well as for the OECD as a whole. Here we analyze the human capital according to three different aspects, each involving several dimensions. The first aspect refers to academic achievements and the situation of the young in the labour market. The second aspect focuses on the situation of the graduates. The third aspect evaluates the academic performance of the 15 year-olds students in the fields of Mathematics and Science, according to PISA 2018.

Section 4 contains a discussion on some of the questions that the empirical application raises (computation, sensitivity analysis, heterogeneous agents and relative valuation). A few final comments on different applications of this methodology closes the work.

2 The evaluation protocol

2.1 The Multidimensional Balanced Worth (MBW)

The reference problem consists of evaluating the relative performance of a set of g populations, $G = \{1, 2, \dots, g\}$, relative to D dimensions. We assume that the achievement in each dimension $d = 1, 2, \dots, D$ is one of the m_d *levels*, ordered from best to worst. Those levels may correspond to categories (qualitative variables) or intervals of a numerical variable.

The ***achievements structure*** of population $i = 1, 2, \dots, g$ can be described by a collection of vectors $\boldsymbol{a}^d(i) = \left(a_{i1}^d, \dots, a_{im_d}^d\right)$, with $d = 1, 2, \dots, D$, where $a_{ir}^d = n_{ir}^d / n_i$ is the fraction of population i that achieves level r in dimension d. Clearly, $a_{ir}^d \geq 0$

and $\sum_{r=1}^{m_d} a_{ir}^d = 1$, for all d.

The basic principle to compare the relative performance of those populations is the probability of getting better outcomes. Let p_{ij}^d denote the probability of an element chosen at random from population i to achieve a higher level than an element chosen at random from population j, both compared concerning dimension d. As the levels are ordered from best to worst, we can calculate that probability as follows:

$$p_{ij}^d = a_{i1}^d\left(a_{j2}^d + \cdots + a_{jm_d}^d\right) + a_{i2}^d\left(a_{j3}^d + \cdots + a_{jm_d}^d\right) + \cdots + a_{i(m_d-1)}^d a_{jm_d}^d$$

Let $e_{ij}^d = e_{ji}^d$ stand for the probability that an observation from i attains the same level as one from j, relative to dimension d. By construction, we have: $p_{ij}^d + p_{ji}^d + e_{ij}^d = 1$. We now define:

$$q_i^d = p_{ij}^d + \left(e_{ij}^d/2\right)$$

as the probability that i beats j in a pairwise comparison regarding dimension d, where we split equally the probability of a tie.

Let us consider first the case in which we have to compare only two populations, i and j. We want to assess the relative performance of those two populations based on the distributions of outcomes across dimensions. Our proposal here is simple and natural: *attach to each population a value that is proportional to the probability of being a winner*.

Bearing in mind that the different dimensions are disjoint, we obtain the expectation that population i beats population j in *some* dimension as:

$$\frac{1}{D}\sum_{d=1}^{D} q_{ij}^d \qquad [1]$$

That is the sum across dimensions of the probabilities that i be the winner over j, divided by the likelihood that each dimension is chosen.

Remark 1: *Note that here we assume that all dimensions are equally likely (or equally important). When this is not the case equation* [1] *becomes a weighted sum,* $\sum_{d=1}^{D} \alpha_d q_{ij}^d$, *with*$\alpha_d \geq 0, \sum_{d=1}^{D} \alpha_d = 1$.

Let us call w_i, w_j the evaluations of populations *i* and *j*, respectively. The proportionality principle stated above implies that:

$$\frac{w_i}{w_j} = \frac{\sum_{d=1}^{D} q_{ij}^d}{\sum_{d=1}^{D} q_{ji}^d} \qquad [2]$$

That is, the ratio of the evaluations is equal to the ratio of the probabilities of one population beating the other in some dimension. Equation [2] has one degree of freedom so that we can choose units arbitrarily. For the case of two populations, therefore, the proportionality principle fully determines the evaluation formula, except for the choice of units.

Equation [2] can be rewritten as:

$$w_i = \frac{\sum_{d=1}^{D} q_{ij}^d w_j}{\sum_{d=1}^{D} q_{ji}^d} \qquad [2']$$

Here the evaluation of group *i* appears as the ratio of two interesting expressions. The one in the numerator corresponds to the *relative advantage* of *i* over *j*, as it gives us the probability of getting better outcomes, weighted by the evaluation of population *j*. The denominator can be interpreted as the *relative disadvantage* of population *i* with respect to *j*, as it expresses the probability of getting worse outcomes.[1]

When there are more than two populations involved, we have to adjust this evaluation protocol, as a chain of pairwise comparisons yields non-transitive outcomes (even in the single-dimensional case). The most natural way to extend previous evaluation avoiding this problem is by taking expectations. That is,

$$w_i = \frac{\sum_{j\neq i}\sum_{d=1}^{D} q_{ij}^d w_j}{\sum_{j\neq i}\sum_{d=1}^{D} q_{ji}^d} \quad , \quad i,j = 1,2,\ldots,g \qquad [3]$$

The numerator now describes the *average relative advantage* of the distribution of population *i* with respect to the rest. In contrast, the denominator corresponds to the *average relative disadvantage* of population *i* with respect to the rest. Trivially, Equation [3] collapses to Equation [2'] when there are only two

[1] One might also consider an evaluation based on the "intersection approach", rather than the "union approach" presented here. That is, calculating the probability that one population beats the other in *all* dimensions (the product of the corresponding probabilities). Yet we consider that the union approach is the natural extension of the evaluation provided by a single dimension. Be as it may the ensuing discussion is compatible with both approaches.

populations.

The vector or those w_i values in equation [3] is called the ***Multidimensional Balanced Worth*** (*MBW* for short), as it consists of an extension of the *balanced worth* in Herrero & Villar (2018) to the case of several dimensions. It provides a *cardinal relative evaluation* of the performance of the different populations concerning several dimensions. That is, it permits one to know not only whether a population performs better than another, but also how much better. It can be shown that the *MBW* vector, $\boldsymbol{w} = (w_1, w_2, \dots, w_g)$, always exists, it is generically unique, except for the choice of units, and positive (see below). It is also interesting to note that it is *monotonic*. That is, if the distribution of population *j* shifts to the upper levels of performance in any dimension, other things equal, the *MBW* of population *j* will increase (a property that implies *stochastic dominance*).

2.3 Existence, uniqueness and positiveness of the MBW

The *MBW* is an intuitive evaluation criterion as it attaches to each population a value that corresponds to the ratio between the average relative advantage and the average relative disadvantage. Let us show that the *MBW* always exists and has the desired properties.

For each dimension *d* = 1, 2, …, *D*, and each population *i*, let:

$$R_i^d = (g-1) - \sum\nolimits_{j \neq i} q_{ji}^d$$

Now consider the following matrix:

$$\mathbf{P}^d = \begin{pmatrix} R_1^d & q_{12}^d & \dots & q_{1g}^d \\ \dots & \dots & \dots & \dots \\ q_{g1}^d & q_{g2}^d & \dots & R_g^d \end{pmatrix}$$

Off-diagonal entries of matrix $\mathbf{P}^d$ correspond to the domination probabilities whereas the diagonal entries reflect the probability of not being dominated, respect to dimension *d*.

We now define the average matrix across all dimensions, as follows:

$$\mathbf{P} = \frac{1}{D}\sum\nolimits_{d=1}^{D} \mathbf{P}^d$$

Matrix **P** summarises the information concerning the evaluation of the *g* populations regarding the D dimensions. By construction, **P** is a square matrix with positive entries (i.e. a Perron matrix) with the property that all its columns add up to $(g-1)$. Therefore (see for instance Berman & Plemmons 1994), **P** has a dominant positive eigenvalue, equal to $(g-1)$, that has associated with it a positive eigenvector, $\boldsymbol{w} = (w_1, w_2, \dots, w_g) \gg \mathbf{0}$, with:

$$\mathbf{P}\boldsymbol{w} = (g-1)\boldsymbol{w}$$

To check that this is the vector we are looking for, note that the ith entry of that eigenvector can be written as:

$$w_i = \sum_{j \neq i} \frac{\sum_{d=1}^{D} q_{ij}^d w_j}{\sum_{d=1}^{D} q_{ji}^d}$$

This eigenvector $\boldsymbol{w}$ is unique and strictly positive, provided that matrix **P** is irreducible. All the entries in matrix **P** are strictly positive, except in the singular case in which for some population the full mass is concentrated at the worst category, and this happens for all the dimensions simultaneously. That is, matrix **P** is generically irreducible.

2.3 An alternative interpretation of the MBW

The *MBW* is an intuitive evaluation criterion that attaches to each population a value that corresponds to the ratio between the average relative advantage and the average relative disadvantage. Let us now provide an alternative interpretation of the *MBW* vector.

Consider the following protocol. We start by choosing at random two populations, i and j, and a dimension d. Next, we calculate the probability that an element from population i achieves a higher level of performance than an element from j, both compared for dimension d. If $p_{ij}^d > p_{ji}^d$, population i is declared the winner in this confrontation; if $p_{ji}^d > p_{ij}^d$, the winner is j. Finally, if $p_{ij}^d = p_{ji}^d$, we toss a coin and luck decides the winner. Now, a new population and a new dimension are randomly chosen to compete with the previous winner. We proceed this way ad infinitum.

This process induces the construction of a matrix **P** that is the transition matrix of a Markov procedure associated with that protocol. The components of the *MBW* vector can be interpreted, therefore, as the proportion of time that each population keeps competing (i.e. is the winner) in the long run, reflecting this way their relative importance.

3 Comparing the human capital of young Europeans

We now apply the multidimensional balanced worth to the analysis of human capital achievements of the young, for a subset of the European countries. Focusing on the young (both teenagers and young adults) amounts to emphasizing the relevance of human capital in the near future. We selected the countries looking for diversity (countries from different European environments) and yet for societies with some common features (all long time members of the European Union at the time), to make the comparison relevant and appealing. The nine selected countries are Finland, France, Germany, Greece, Ireland, Italy, Netherlands, Spain and the United Kingdom (UK). We also include the average values of the OECD countries as a benchmark.

Human capital is a multifarious concept that refers to some critical features of society that affect development possibilities and social welfare. We do not pretend here to establish how it should be measured but simply propose considering three different aspects worth studying. First, the educational achievements and the situation for the studies and the labour market of young adults. Second, the situation of those young adults with tertiary education regarding the labour market and intergenerational mobility. And third, the quality of compulsory schooling as approximated by PISA results in the fields of mathematics and science. Each of these aspects involves several dimensions.

All tables that contain primary data have the OECD as source ("Education at a Glance", from 2017 and 2019, and the 2018 PISA report). Those tables describe the distributions of the corresponding populations in categorical variables, under the implicit assumption that columns in the tables are arranged from best (left) to worst (right).

3.1 Educational achievements and the labour market.

The first aspect of human capital we address here refers to the educational achievements and the situation of the young regarding studies and the labour market. More precisely, we shall focus on these two dimensions:

(1) *Educational achievements*. We chose a variable that aims at approaching the formal education attained by young adults (25-34 year-olds). Here we consider the distribution of the population into three different educational levels: Tertiary education, upper secondary education, and less than upper secondary education.

(2) *Education and the labour market*. Here we concentrate on the population between 18 and 24 years old. We look at their situation in the educational system and the labour market (in particular those who are enrolled in formal education or are already employed, those who are unemployed, and those who belong to neither of those categories).

Tables 1 and 2 present the primary data. Table 1 refers to educational achievements. It shows that there is some diversity among European young adults regarding those achievements. Finland and Ireland have over 90% of that population with at least upper secondary education. Italy and Spain have much smaller values but are also quite different regarding the distribution between tertiary studies and upper secondary studies. Germany also exhibits a relatively small fraction of that population with tertiary studies (e.g. as compared with the UK or even the OECD average).

Table 1: Educational attainment 25-34 year-olds, 2018

	Tertiary	Upper Secondary	Less than Upper Secondary
Finland	41	49	10
France	47	40	13
Germany	32	55	13
Greece	43	44	13
Ireland	56	36	8
Italy	28	48	24
Netherlands	48	40	12
Spain	44	23	33
UK	51	34	15
OECD av.	44	41	15

Source: OECD 2019a

Table 2 conveys information on the situation of the young regarding their activity in the educational system and the labour market. We consider here three different situations: (1) Those who are studying or employed; (2) Those who are neither studying nor employed but are looking for a job (hence, still part of the active population); and (3) Those who neither study nor belong to the active population, usually known as NEET, the acronym of "Not in Education, Employment or Training". Here we find once more very diverse situations. Regarding the shares of the young in education or employed, we observe that all selected countries are above the OECD average, except for Greece, Italy and Spain. If we focus on the unemployed or those that left both their studies and the labour force, we find the worst values for Greece (9 points below the OECD mean), Italy (11 point below) and Spain (6 points below). Especially worrying is the case of Italy with more than 14% of the young in the category of NEET.

Table 2: Percentages of 18-24 year-olds in Education and the labour market, 2018

	In Education or Employed	Unemployed	NEET
Finland	**87**	6	7,1
France	**88**	3,8	7,4
Germany	**90**	3,1	6,5
Greece	**78**	14	8,2
Ireland	**88**	5	7,6
Italy	**75**	11,5	14,3
Netherlands	**93**	1,6	5,3
Spain	**80**	12	8,2
UK	**87**	5	8,8
OECD av.	**86**	5,7	8,6

Source: OECD 2019a

How can we integrate those two dimensions into a single measure, without renouncing to use all available information regarding the distributions in the different categories? The MBW is a sound way of dealing with this evaluation problem and keeping track of that information.

Table A provides the comparative evaluation in terms of the multidimensional balanced worth. We also include the values corresponding to the individual balanced worth of each of the two dimensions considered (we follow the ordering of the above tables). In all cases, we normalise the vectors by letting the value of the OECD be equal to 1.

The evaluation shows that, overall, Ireland and the Netherlands are well above the OECD average (Ireland doing especially well concerning educational achievements). Italy, on the other extreme, is more than 25% below the OECD benchmark, with values particularly low in educational achievements. Finland, France and the UK are slightly above the OECD mean whereas Germany, Greece and Spain are below, with different impacts of both dimensions.

Table A: Multidimensional balanced worth and balanced worth of the different dimensions

	MBW	BW1	BW2
Finland	1,016	1,005	1,026
France	1,068	1,077	1,059
Germany	0,956	0,833	1,095
Greece	0,933	1,005	0,865
Ireland	1,173	1,331	1,036
Italy	0,739	0,684	0,796
Netherlands	1,132	1,108	1,157
Spain	0,856	0,816	0,898
UK	1,067	1,127	1,010
OECD av.	**1**	1	1

3.2 The graduates

The second aspect we consider refers to those young adults with tertiary studies. This collective deserves special attention at least for two reasons. First, because it is the population in which society has invested more in human capital. Second, since that part of the population has more chances to adapt and lead economic and social changes.

We also consider here two different dimensions:

(i) *Employment*. Here we observe the employment rates of young adults with tertiary studies, aged between 25 and 34 years.

(ii) *Intergenerational mobility*. We measure this dimension through the shares of the population of adults between 30 and 44 years-old with tertiary education, whose parents did not reach that level of education (a resilience feature).

Table 3 reports the employment rates of those young adults with tertiary education. Those rates are very high in general (84% for the OECD as a whole), even though there are substantial differences (92% employed in the Netherlands versus 67% in Italy).

Table 3: Employment rates of 25-34 year-olds with tertiary education, 2018

	Employment rate (%)	Unemployment (%)
Finland	85	15
France	83	17
Germany	88	12
Greece	70	30
Ireland	87	13
Italy	67	33
Netherlands	92	8
Spain	78	22
UK	90	10
OECD av.	84	16

Source: OECD 2019a

Table 4 describes the distributions of young adults whose parents did not reach tertiary education into three different categories: those who did not reach tertiary education, those who finished undergraduate studies and stopped there (Tertiary B), and those who did some graduate studies (master or doctorate, called here Tertiary A).

Focusing on the third column, we observe that Finland, Italy, UK and Ireland present values of intergenerational mobility clearly above the mean of the OECD. France, Netherlands and Spain are around that mean. Germany and Greece exhibit values well below that reference figure. As for the shares of those with graduate studies (Tertiary A), Finland, Netherlands, Italy and the UK have the highest scores whereas France, Germany and Greece obtain the lowest values.

Table 4: Tertiary attainment 30-44 year-olds whose parents both have less than tertiary education, 2015

	Tertiary A	Tertiary B	No Tertiary
Finland	32	15	52
France	16	15	69
Germany	14	11	75
Greece	14	10	76
Ireland	19	16	65
Italy	27	14	59
Netherlands	28	4	68
Spain	20	12	68
UK	25	13	62
OECD av.	20	12	69

Source: OECD 2017

The first column of the body in Table B shows the values of the multidimensional balanced worth. Finland, the Netherlands and the UK are clearly above the OECD average. Ireland is slightly above that mean value whereas France, Germany, Spain, Italy and Greece are below the OECD benchmark. We also include, as before, the balanced worth calculated independently for each dimension. Regarding employment rates (BW3), the Netherlands and the UK are clearly above the OECD reference; Italy, Greece and Spain present the lowest values. Finland and Italy have the highest scores relative to intergenerational mobility (BW4), whereas Greece, Germany and France show poor results.

Table B: The MBW for Graduates

	MBW	BW3	BW4
Finland	1,192	1,020	1,396
France	0,972	0,980	0,964
Germany	0,969	1,084	0,865
Greece	0,801	0,753	0,852
Ireland	1,055	1,062	1,048
Italy	0,930	0,708	1,217
Netherlands	1,117	1,175	1,064
Spain	0,945	0,886	1,007
UK	1,136	1,129	1,145
OECD av.	1	1	1

3.3 The quality of compulsory education in maths and science

Here we propose to use PISA data in the fields of mathematics and science to assess the achievements of the 15-year old students for the same group of countries. As 16 is the age at which compulsory education ends in those countries, we can regard PISA data as a proxy of the knowledge that each society guarantees to its citizens. In that sense, we speak of the quality of compulsory education.

Tables 5 and 6 show the distributions of the 15-year old students in five levels of proficiency regarding the outcomes of the last wave of the PISA test, in the fields of mathematics and science. The reason why we have chosen those two fields, skipping reading comprehension (the main topic of the 2018 PISA wave), is twofold. First and foremost, because there was a problem with the results of reading comprehension in one of the selected countries (Spain), for which there are no data. And second, because Mathematics and Science are part of the so-called STEM subjects, which seem to have particular relevance for the near future.

Those proficiency levels define differential abilities of the students and are parameterized by intervals of the test scores (see OECD 2019b for details). Let us recall here that there is a well-established convention according to which those students below level 2 are regarded as low-performers. In contrast, those within levels 5 and 6 are high-performers. We have included in the tables the coefficients of variation (CV) of the nine selected countries, to emphasize how the more significant differences between countries achievements appear precisely in the tails (high and low performers).

Regarding mathematics, the share of high performers in the Netherlands is much larger than that of the OECD. Germany, the United Kingdom, Finland and France are also above the mean. Greece has a tiny share of high performers, as it is also the case (even though not so much) of Spain and Ireland. Greece is also the country with the worst values regarding low performance, followed by Spain (the only two countries in our sample with values worse than those of the OECD).

Table 5: Percentages of students at each proficiency level. Mathematics. PISA 2018

	Levels 5 & 6	Level 4	Level 3	Level 2	Level 1 and below
Finland	11,13	22,67	28,94	22,30	14,98
France	11,02	21,04	25,58	21,10	21,26
Germany	13,32	20,82	24,03	20,74	21,10
Greece	3,74	11,10	22,49	26,83	35,84
Ireland	8,23	20,83	30,53	24,72	15,69
Italy	9,54	18,12	25,61	22,91	23,82
Netherlands	18,42	23,58	23,22	19,03	15,75
Spain	7,28	17,54	26,03	24,45	24,70
United Kingdom	12,86	20,43	25,46	22,02	19,23
OECD average	10,93	18,53	24,36	22,21	23,98
CV nine countries	0,370	0,178	0,094	0,098	0,286

Source: OECD 2019b

As for the field of science, the data show a similar pattern with some differences. Finland is the country with the highest share of high performers, followed by the Netherlands, Germany and the UK. Spain, Italy and Greece present the lowest percentages. Only Italy and Greece exhibit shares of low performers higher than de OECD average, followed by Spain, the Netherlands and France.

Table 6: Percentages of students at each proficiency level. Science. PISA 2018

	Levels 5 & 6	Level 4	Level 3	Level 2	Level 1 and below
Finland	12,3	24,9	28,9	21,1	12,9
France	6,5	20	28,3	24,6	20,5
Germany	10	21,5	26,9	22	19,6
Greece	1,3	9,3	26	31,6	31,7
Ireland	5,9	19	31,3	26,9	17
Italy	2,8	13,4	27,8	30,2	25,9
Netherlands	10,6	22,1	24,9	22,4	20
Spain	4,2	16,8	29,4	28,4	21,3
United Kingdom	9,7	20,8	28,1	24	17,4
OECD average	6,7	18,1	27,4	25,8	22
CV nine countries	0,512	0,243	0,064	0,138	0,247

Source: OECD 2019b

Table C provides the multidimensional balanced worth of this aspect and the corresponding balanced worth for maths (BW5) and science (BW6). Altogether Finland and the Netherlands exhibit values well above the mean of the OECD, with Finland relatively much better in science and the Netherlands in maths. UK and Germany also have values above the mean (even France, but much closer to it). Spain is also close to the OECD but from the other side of the mean, whereas Italy and especially Greece present much lower values.

Table C: The MBW (and BW) for PISA in Maths and Science (2018)

	MBW	BW5	BW6
Finland	1,343	1,230	1,468
France	1,074	1,089	1,059
Germany	1,159	1,134	1,186
Greece	0,643	0,635	0,651
Ireland	1,102	1,115	1,089
Italy	0,879	0,971	0,795
Netherlands	1,296	1,414	1,188
Spain	0,921	0,904	0,939
United Kingdom	1,173	1,147	1,199
OECD average	1	1	1

4 Discussion

We have presented a procedure to evaluate the relative performance of different populations in a multidimensional context when performance is described by categorical data. This procedure exploits in full the information regarding the distributions of the populations into the corresponding levels of performance, rather than comparing average values. The evaluation protocol, named multidimensional balanced worth, provides a quantitative assessment of such performance based on an easy and intuitive formula: the ratios of average advantages and disadvantages, duly computed. Such a method derives from an old friend in this endeavour: the stable distribution of a Markov chain. Which is interesting in a twofold way. First, because it is a well-known mathematical tool so that we understand quite well what it does and how it works. Second, because it makes calculations conventional.

We have also illustrated the working of this evaluation protocol by measuring three different aspects of human capital in the young, for nine selected European countries, with each of those aspects consisting of two dimensions. Let us now discuss some elements that arise from the application of this evaluation method in empirical studies.

4.1 Computation, weighting and sensitivity analysis

Regarding computational complexity, let us point out from the outset that we have developed a friendly and freely accessible algorithm that can be used to perform calculations by just plugging the matrices that describe de distributions of the populations in the corresponding categories (raw data). The algorithm builds up internally the associated $\mathbf{P}^d$ matrices, the summary $\mathbf{P}$ matrix, and then calculates the dominant eigenvector normalised by the mean.[2]

One of the advantages of this computational facility is that it permits one performing sensitivity analysis at practically no cost. This fact might be essential to assess the relevance of the number of categories and dimensions in the overall

[2] Thanks are due to Héctor García Peris (Ivie) for programming this algorithm, which is available at the Ivie web site, in: http://web2011.ivie.es/multidimensional-balanced-worth/

evaluation and permits one to add or remove dimensions easily. The algorithm also allows us to impute different weights to different dimensions, which makes the sensitivity analysis still more interesting.

Let us illustrate those aspects by integrating the employment rate of the young with tertiary studies (see Table 3) as a new dimension in the analysis of section 3.1 regarding educational achievements and the labour market. So now we consider that the analysis of educational achievements and the labour market involves three dimensions, rather than two. As those with tertiary education represent some 44% of the total of the young in the OECD, we can weigh this dimension accordingly, so that in the new evaluation we have three dimensions with weights $\left(\frac{1}{3}+\frac{1}{2}\left(0.66\frac{1}{3}\right)\right)$ for the first two, and $\left(0.44\frac{1}{3}\right)$ for the third one. Applying those adjustments, we obtain the multidimensional balanced worth (MBW+) that appears in Table D. The table also shows the relative differences between this new appraisal and the original one in section 3.1. Including the employment rates of the graduates' changes very little the outcomes. It reduces the values of France, Greece and Ireland slightly and introduces small improvements in the rest.

Table D: The multidimensional balanced worth in the extended case

	MBW+	Relative difference
Finland	1,017	0,07%
France	1,055	-1,24%
Germany	0,974	1,87%
Greece	0,904	-3,05%
Ireland	1,156	-1,44%
Italy	0,734	-0,60%
Netherlands	1,138	0,56%
Spain	0,860	0,53%
UK	1,076	0,84%
OECD av.	1	0

Let us remark that the data regarding the graduates in Table 3 are simply the percentage of those who are employed, which we have converted into distribution by just adding its complement. This shows that we can also apply our approach in a context in which the available information is poor. In any case, the richer the information we get, the better the approximation we obtain. This remark only intends to point out that if there is a dimension deemed relevant on which we have little information, we can still incorporate it into the evaluation (bearing in mind that we only have a rough estimate of its impact).

4.2 Heterogeneous populations

Let us now refer to the case of heterogeneous populations. It is implicit in our evaluation process that all societies compared consist of homogeneous agents. There are many instances, however, in which there is heterogeneity in the population, due to age, gender, race, wealth, location, etc., which may be relevant for the evaluation. Acknowledging heterogeneity amounts to consider that each population consists of different *types*. As described in detail in Herrero & Villar (2018), there are several different questions one can address in this richer context.

We illustrate here this aspect by revisiting once more the evaluation presented in section 3.1, in which we measure human capital in terms of two dimensions, educational achievements and education and the labour market. Now the population of each country is considered as composed of two different types, women and men. We can, therefore, think of an evaluation problem involving 2×10 population subgroups and carry out a *joint evaluation*. That is, it implies that we think it relevant comparing the human capital of German women with that of Italian men, say. Another possibility is that of making comparisons only among the same population subgroups (e.g. human capital of the young European women). We call this sort of comparison *separate evaluation by types*, as it provides an assessment of the *between populations* relative performance by types. A third possibility is comparing population subgroups within countries (e.g. human capital between young women and men in France). We call this type of comparison *separate evaluation by populations*, as it provides a measure of *within populations* heterogeneity.

Each one of the previous comparisons provides an evaluation from a different angle, and which one is better fit depends on the focus of our analysis.

Tables 7 and 8 provide the basic information on the distribution of the populations, divided by gender, within the different categories.

Table 7: Educational attainment 25-34 year-olds, by gender, 2018 (%)

	Men			Women		
	Tertiary	Upper Sec	Below US	Tertiary	Upper Sec	Below US
Finland	34	55	11	50	42	8
France	43	43	14	51	37	12
Germany	31	55	14	34	54	12
Greece	35	50	15	51	38	11
Ireland	52	39	9	60	34	6
Italy	22	51	27	34	45	21
Netherlands	43	42	15	52	37	11
Spain	38	24	38	50	23	27
UK	48	35	17	54	33	13
OECD av.	38	46	17	51	36	13

Source: OECD 2019a

Table 8: Percentages of 18-24 year-olds in Education and the labour market, by gender, 2018 (%)

	Women			Men		
	In Ed+Em	*Unem*	*NEETS*	*In Ed+Em*	*Unemp*	*NEETS*
Finland	85,98	5,14	8,88	83,96	8,49	7,55
France	82,18	10,39	7,43	80,43	12,49	7,08
Germany	90	2,73	7,27	91,00	3,90	5,10
Greece	75,77	15,81	8,42	78,15	13,49	8,36
Ireland	87	5,76	7,24	87,00	7,28	5,72
Italy	73,98	11,37	14,65	72,91	13,85	13,24
Netherlands	93,29	1,65	5,06	92,47	2,22	5,31
Spain	79,86	11,81	8,33	78,40	14,82	6,78
UK	86,14	3,60	10,26	86,63	6,76	6,61
OECD av.	84,49	5,09	10,42	86,50	6,97	6,53

Source: OECD 2019a

Tables E, F and G provide the different types of evaluation when we take into account that societies consist of women and men.

Table E presents the separate evaluation by gender. Comparisons only make sense along with columns, that is, comparing the human capital of women (resp. men) for the selected countries. Women in Finland, Ireland, Netherlands and UK are above the OECD women's average. Men in France, Germany, Ireland, Netherlands and the UK are above the men in the OECD.

Table E: MBW Men/Women. Separate evaluation by gender

	Men	Women
Finland	0,981	1,034
France	1,005	0,989
Germany	1,007	0,916
Greece	0,913	0,936
Ireland	1,191	1,153
Italy	0,717	0,745
Netherlands	1,121	1,114
Spain	0,816	0,887
UK	1,090	1,042
OECD av	1	1

Table F provides the separate evaluation by populations. Here we normalize the human capital of men setting its value equal to 1 in each country. The comparison makes only sense along rows, and each value of the right column can be regarded as the percentage of the human capital of women relative to men. The most outstanding fact is that all countries (and the OECD as a whole) show that women have better human capital than men, particularly in Greece, Italy and Spain.

Table F: MBW separate evaluation by countries

	Men	Women
Finland	1	1,089
France	1	1,101
Germany	1	1,029
Greece	1	1,151
Ireland	1	1,089
Italy	1	1,150
Netherlands	1	1,113
Spain	1	1,166
UK	1	1,063
OECD av.	1	1,119

Finally, Table G shows the joint evaluation. Here we compare men and women of different countries concerning the average human capital of the young males in the OECD. That means, for instance, that French women have some 11%

higher human capital than the average OECD male, whereas French males are very much around that average value. Note that women of all selected countries, other than those from Italy and Spain, exhibit higher human capital than the average young male in the OECD, with the highest values corresponding to Ireland and the Netherlands (actually the absolute highest one).

Table G: MBW joint evaluation by country and gender

	Men	Women
Finland	0,978	1,159
France	1,005	1,106
Germany	1,004	1,029
Greece	0,912	1,046
Ireland	1,189	1,293
Italy	0,719	0,832
Netherlands	1,121	1,248
Spain	0,826	0,985
UK	1,093	1,165
OECD av	1	1,118

4.3 All is relative

This evaluation method is based on the comparison of the shares of the populations into different performance levels. It is a relative evaluation in a twofold way. On the one hand, it is relative because the value attached to each society is a function of the values assigned to the remaining (in the wording of social choice theory, there is no independence of irrelevant alternatives here). On the other hand, because if we increase the size of any population by replicating its members several times, the evaluation will not change as the shares would remain unaltered (so here the evaluation is relative in the sense that disregards absolute values). This second feature, usually called replication invariance, is important because it permits one to perform comparisons of populations of different size. Yet in some cases, one might be interested in keeping track of the relative differences in size. The extension of the balanced worth to the multidimensional case may help in this task, as illustrated next.

Let us consider the evaluation of intergenerational mobility discussed in section 3.2 as a single-dimensional evaluation problem on its own. Table 4 describes the distributions of young adults whose parents did not reach tertiary education into three different categories. One may be willing to compute not only those shares but also the relative size of this group in each country. That is the percentage of the young, aged between 30 and 44, whose parents lack tertiary education. The idea is that the larger this percentage, the more important is each percentual point of those who achieve tertiary studies (a more widespread resilience).

Table 9 provides these data and shows that there are relevant differences among the countries. In Italy, that figure reaches 95%, followed by Spain and Greece, with 88% and 87%, respectively. Only Germany, the UK and the Netherlands have values below the OECD average.

Table 9: Distribution of the young aged 30-44 depending on whether their parents both have less than tertiary education, 2015 (%)

	Without TE	With TE
Finland	78	22
France	80	20
Germany	65	35
Greece	87	13
Ireland	78	22
Italy	95	5
Netherlands	73	27
Spain	88	12
UK	71	29
OECD av.	75	25

Source: OECD 2017

We can now consider the multidimensional evaluation problem that obtains when we consider the proportion of those young as a new dimension. That is, taking the data in Table 4 and Table 9 as two dimensions of the same evaluation problem. Table H provides the results under the assumption that both dimensions have equal weights.[3]

The first column of Table H shows the values of the multidimensional balanced worth. All of the selected countries, but Germany, are above the OECD average. Italy, Finland and Spain exhibit the highest values.

[3] In this case we consider the distribution young whose parents have no tertiary education as a "better" category than its complement. This is not because we think that it is actually better to have parents without tertiary education but because it is a positive indicator of social mobility.

Table H: The MBW for intergenerational mobility

	MBW	BW4	BW9
Finland	1,216	1,396	1,062
France	1,033	0,964	1,106
Germany	0,841	0,865	0,816
Greece	1,041	0,852	1,274
Ireland	1,055	1,048	1,062
Italy	1,350	1,217	1,500
Netherlands	1,011	1,064	0,960
Spain	1,143	1,007	1,300
UK	1,028	1,145	0,922
OECD av.	1	1	1

This is just an example to illustrate the argument using the data already discussed in Section 3. There are other cases in which this aspect may be much more relevant. Let us mention two of them involving participation rates.

One of the problems that we face when comparing PISA data of countries with different levels of development is the variety of participation rates. PISA scores refer to those 15-year old students who attend school. There are countries, even within the OECD, in which a substantial part of the 15-year-old does not attend school. That implies that comparing distributions by levels of performance introduces a bias. Including participation rates as a new dimension may help to get better estimates of the differences between educational systems in a consistent and non-arbitrary way.

A similar problem arises when comparing the distribution of unemployment between countries in terms of duration.[4] Here again, we can find substantial differences in participation rates between countries that affect the interpretation of the comparison of unemployment structures. By introducing the distribution of the working-age population between active and inactive as a new dimension allows computing systematically those differences.

[4] The standard categories are: unemployed for less than one month, between one and three months, between three and six months, between six months and one year, between one year and two years and more than two years

5 Final comments

The evaluation protocol proposed here provides a tool that applies to multidimensional problems that involve categorical data. It is flexible, intuitive and computationally costless. As it is an extension of the previous methods (worth and balanced worth), it is interesting to conclude by mentioning some of the applications that have used these methods for single-dimensional problems, and see how they can be extended to richer scenarios.

The original paper (Herrero & Villar 2013) includes three applications to this family of problems. The first one refers to the composition of human capital in different European countries by comparing the distribution of the working-age population across educational attainments. This topic has been recovered in this paper and extended into a multidimensional scenario in different directions (Section 3). The second application refers to the evaluation of the cognitive abilities of the adult population in some OECD countries, using the data of the 2013 PIAAC report, concerning reading literacy. We have also tackled here a similar problem in Section 3.3 using PISA data regarding mathematics and science (see also the related contributions in Herrero, Méndez & Villar 2014 and Villar 2014). Their last application provides a comparative evaluation of the perceived health situation of the EU15 countries, taken from the Self-reported Health Status Survey. This application is interesting because it shows a way of dealing with problems in which the informational inputs are subjective perceptions and categories are purely ordinal. A multidimensional health indicator that combines perceived health with other objective variables is easy to design.

Gallen and Peraita (2015) provide an application of the worth to the analysis of corporate social responsibility (CSR) engagement in the OECD. The interest of this question derives from the observed expansion of CSR engagement of the OECD countries in recent years, a period of the financial crisis. Torregrosa (2015) uses the worth to analyse the evolution of autonomic-nationalist feelings in Spain based on opinion surveys regarding the state of Spanish Autonomous Communities carried out by Spain's Centre for Sociological Research since 1996. Extensions of these applications are also natural.

Albarrán *et al.* (2017) analyse the intellectual influence in terms of the citation impact of published research. The analysis is based on a dataset consisting of 4.4 million articles published in the period 1998-2003 and indexed by Thomson Scientific, as well as the citations they received during a five-year citation window for each year in that period. Countries are compared by research areas. A natural multidimensional extension of this study would be to compare countries simultaneously in all areas. Considering each research area a different dimension and recurring to the multidimensional balanced worth is the way of so doing.

Herrero & Villar (2018) provide an application on life satisfaction in Spain by age groups and gender, using a qualitative, single-dimensional self-reported variable. A multidimensional version of this evaluation might make much more sense (think for instance of the OECD *How's Life* programme).

Fernandez-Herrero & Lorenzo-Lledó (2019) analyse the effects of using a special virtual reality tool on the improvement of children with autism in social engagement, where their performance in different aspects are categorical. They consider four separate items: Social and emotional reciprocity; non-verbal communication; inflexibility to changes, and stereotypes and sensorial reactivity. The MBW may be used to evaluate all aspects simultaneously.

Pita & Torregrosa (2020) applied this methodology to analyse the gender-job satisfaction paradox by using several waves of the European Working Conditions Survey. Out of 106 questions, they focused on Q88. Indeed, there is plenty of room to incorporate additional questions to obtain more robust results by using the MBW.

Finally, let us mention the work in Herrero & Villar (2020) applying this evaluation protocol to the analysis of income opportunities, using a novel way of approaching the study of income distributions. Here again, adding dimensions may provide a richer picture of societies (e.g. including health and educational variables, in the spirit of the United Nations human development index).

References

Albarrán, P., C. Herrero, J. Ruiz-Castillo & A. Villar (2017), The Herrero-Villar approach to citation impact, **Journal of Informetrics**, 11(2): 625-640.

Berman, A. & Plemmons, R. J. (1994), **Nonnegative matrices in the mathematical sciences**. Society for Industrial and Applied Mathematics.

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. **Computer Networks and ISND systems,** 30: 107-117

Boccard, N. (2020), Pairwise Comparison Ranking. The Landau Method, mimeo, University of Girona.

Daniels, H. E. (1969), Round-robin Tournament Scores, **Biometrika**, 56: 295-299.

Fernandez-Herrero, J & Lledo, L. (2019), An immersive virtual reality educational intervention on people with autism spectrum disorders (ASD) for the development of communication skills and problem-solving**, Education and Information technology**, 1-34. doi.org/10.1007/s10639-019-10050-0

Freeman, L.C. (1977), A set of measures of centrality based on betweenness, **Sociometry**, 40: 35-41.

Gallén, M.L. and Peraita, C. (2015), A comparison of corporate social responsibility engagement in the OECD countries with categorical data, **Applied Economics Letters**, 22: 1005-1009.

Herrero, C., Méndez, I. and 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. and 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. & Villar, A. (2018), The Balanced Worth: A procedure to evaluate performance in terms of ordered attributes, **Social Indicators Research**, 140: 1279-1300.

Herrero, C. & Villar, A. (2020), Opportunity Advantage. A New Approach to Comparing Income Distributions, Ecineq working paper 2020 516.

Keener, J.P. (1993), The Perron-Frobenius Theorem and the ranking of football teams, **SIAM Review**, 5(1): 80-93.

Laslier, J. (1997), **Tournament solutions and majority voting.** Springer, Berlin, Heidelberg, New York.

Liebowitz, S., J. Palmer (1984), Assessing the relative impact of Economic Journals, **Journal of Economic Literature**, 22: 77-88.

Moon, J.W., N.J. Pullman, (1970), On Generalized Tournament Matrices, **SIAM Review**, 12(3): 384-399.

Newmann, MEJ, (2003), The structure and function of complex networks, **SIAM Review**, 45: 167-256.

OECD (2017), **Education at a Glance 2017: OECD Indicators**, OECD Publishing, Paris.

OECD (2019a), **Education at a Glance 2019: OECD Indicators**, OECD Publishing, Paris.

OECD (2019b), **PISA 2018. Results. What Students know and can do (Vol. I),** OECD Publishing, Paris

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

Pifarré i Arolas, H., Dudel, C. (2019), An Ordinal Measure of Population Health, **Social Indicators Research**, 143 (3): 1219-1243.

Pinsky, G., F. Narin (1976), Citation Influence for Journal Aggregates of Scientific Publications: Theory, with applications to Physics, **Information Processing, and Management**, 12: 297-312.

Pita, C; Torregrosa, R. (2020), The Gender-Job Satisfaction Paradox through Time and Countries, Mimeo, University of Salamanca. Doi: 10.13140/RG.2.2.12266.67529

Saaty, T.H. (2003), Decision-making with the AHP: Why is the principal eigenvector necessary, **European Journal of Operational Research**, 145: 85-91.

Slutzki, G, O. Volij, (2005), Ranking participants in generalized tournaments, **International Journal of Game Theory**, 23(2): 255-270.

Slutzki, G, O. Volij, (2006), Scoring of web pages and tournaments: axiomatizations,

Social Choice and Welfare, 26 (1): 75-92.

Torregrosa, R. (2015), Medición y evolución del sentimiento autonómico en España, **Investigaciones Regionales/Journal of Regional Research**, 33 : 53-70.

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**, vol. 1, nº 1, pp. 72-88.

Wasserman, S., K. Faust (1994), **Social Network Analysis: Methods and Applications.** Cambridge, Cambridge University Press.

serie ad

Ivie

Guardia Civil, 22 - Esc. 2, 1º
46020 Valencia - Spain
Phone: +34 963 190 050
Fax: +34 963 190 055

Website: www.ivie.es
E-mail: publicaciones@ivie.es