World State of Quality: a frontier approach to benchmark the performance of countries worldwide

Vilarinho, H.[1], Cubo, C.[2], Sampaio, P.[2], Saraiva, P.[3,4], Reis, M.[3], Nóvoa, H.[1] and Camanho, A.S.[1]

[1] Faculdade de Engenharia, Universidade do Porto, Porto, Portugal.

[2] Centro de Investigação ALGORITMI, Escola de Engenharia, Universidade do Minho, Braga, Portugal.

[3] Universidade de Coimbra, CIEPQPF, Departmento de Engenharia Química, Coimbra, Portugal.

[4] NOVA Information Management School (NOVA IMS), Universidade Nova de Lisboa, Lisboa, Portugal.

STRUCTURED ABSTRACT

Purpose - The World State of Quality (WSQ) Project aims to evaluate, analyse, rank and categorise countries according to their performance in quality as a multidimensional concept. The Project involves the computation of an overall score for each country, obtained as a weighted average of ranking positions of 16 metrics, with weights determined by a panel of experts.

Methodology- This work proposes an alternative strategy for that procedure, using a Benefit-of-the-Doubt (BoD) Composite Indicator approach under the framework of Data Envelopment Analysis (DEA). This strategy avoids the need of using subjective weights and normalising data by rank positions, using a more objective procedure to obtain the countries' ranking. A new overall score of the World State of Quality is proposed, which allows the categorisation of countries' performance. The novel insights resulting from the use of this methodology are discussed, including the identification of strengths and weaknesses of the various countries, and the peers that can be used for facilitating continuous improvements policies.

Findings - The results show that the BoD approach and the original method used by the WSQ Project present comparable results. Countries' strengths and weaknesses and their suitable peers and targets for benchmarking are presented with illustrative examples.

Originality/value – A novel frontier approach for countries' benchmarking regarding their performance in quality is proposed, incorporating new insights into the current method.

Keywords: macroquality, composite indicator, benchmarking, Benefit-of-the-Doubt.

Paper type: Research paper

1 INTRODUCTION

The comparison of companies or business units regarding their performance in quality is a well-established procedure, often employed to identify best practices and foster continuous improvement. However, assessing quality performance at a country level is a challenging task, requiring a broader approach to deal with different contexts, cultures and societal environments. This macroquality concept was discussed by Saraiva et al. (2020) and used as a foundation to develop the World State of Quality (WSQ) Project, aiming to evaluate countries regarding their quality performance. In this context, a comparative analysis of countries can support the identification of priorities for improvement and development of the best policies to enhance national quality levels. The macroquality assessment includes enablers and activities specific to several quality dimensions and considers outcomes related to the achievement of a sustainable and cohesive environment.

The WSQ Project was developed by a research team from Portugal (Cubo et al., 2019; Sampaio et al., 2018; Saraiva et al., 2020), and the initial findings were published using data collected in 2016 from a set of public databases regarding the performance of European countries. Since 2017, the team extended the analysis to countries outside Europe, such that the Project could achieve a worldwide scope. In this context, the Overall World State of Quality Ranking Score (OWQS) is computed for each country based on 16 indicators, considering ten different dimensions. This overall score is obtained from a weighted average of the indicators' rank positions, where the weights are defined by a group of quality experts. Based on the final ranking according to OWQS values, the countries are classified into four categories.

In this work, an alternative strategy to the estimation of the OWSQ, the construction of a Composite Indicator (CIs) based on the "Benefit-of-the-Doubt" (BoD) modelling approach is proposed. The BoD models are built under the Data Envelopment Analysis (DEA) framework and have been broadly adopted as benchmarking tools (Nardo et al., 2008). This approach can bring new insights compared to the current method for a variety of reasons. First of all, BoD models do not require the prior assignment of aggregation weights, as the weights are endogenous to the existing data and estimated using optimisation techniques. These models can deal with data in their original form, so they do not need to be transformed using rank positions. Furthermore, the models enable the identification of the strengths and weaknesses of the decision-making units (DMUs), as well as the selection of the most suitable peers and targets for benchmarking purposes. Those novel insights represent the literature gap that this work intends to fill. The objectives of this work are as follows: (i) to rank countries based on their macroquality performance using a new method, (ii) to propose a new classification of countries into four WSQ categories, (iii) to identify countries' strengths and weaknesses, and (iv) to determine appropriate targets and benchmarking peers.

This paper unfolds as follows. A brief literature review, covering country evaluations regarding quality performance and the construction of CIs using a BoD approach is presented in Section 2. Section 3 introduces the dataset utilised for the analysis. The proposed methodology is described in Section 4. The results are discussed in section 5, and section 6 presents the conclusions.

2 LITERATURE REVIEW

The literature review detailed in the following subsections covers different types of studies comparing countries regarding their performance in quality. The use of BoD models to produce composite indicators (CIs) for benchmarking is also discussed.

2.1 Quality performance assessment of countries

Comparative studies of country performance using CIs are frequently used to identify strengths and weaknesses and estimate the potential for improvement. Some of the topics that received special attention in the literature include competitiveness (Aiginger, 1998), innovation (Dutta et al., 2020), ease of doing business (Qazi et al., 2021; Rogge and Archer, 2021), inequality (Farris, 2010) and happiness (Helliwell et al., 2020).

Assessments and comparative studies of quality practices are usually performed at an organisational level. They are often based on quality system models, such as the renowned ISO 9000 set of standards (Ahmed et al., 2005; Ismail et al. 1998; Chen et al., 2019). A few studies are also based on the implementation of diverse quality management frameworks, such as Six Sigma (Van Iwaarden et al., 2008), or criteria outlined in quality award models, as the European Foundation for Quality Management (EFQM) Excellence Model (Bou-Llusar et al., 2009) and the Malcolm Baldridge National Quality Award (Jaeger et al., 2013; Lau et al., 2004).

Comparative studies have also been performed involving firms in different countries, such as Turkey and United Kingdom (Clegg et al., 2013), Argentina and Uruguay (Bello-Pintado and Merino-Díaz-de-Cerio, 2013) or comparing Eastern and Western nations (Dahlgaard et al. 1998). Two global studies comparing more than a thousand companies worldwide covering themes such as customer impact, industry standards' trends, training methods and governance structures were published by the American Society of Quality (APQC and ASQ 2013, 2016). Those researches also highlight similarities and contrasts between countries and regions, but all of them focus on the comparison of firms in different nations as a way to assess the diverse scenarios of quality practice implementation. Nevertheless, studies comparing countries' performance in quality at a macro level are scarce.

Determining quality or excellence at the regional or national level, rather than at the level of a single business, is a difficult task. According to Saraiva et al. (2020), defining quality on a larger scale requires a multidimensional approach called "macroquality". This notion encompasses not only quality related actions and facilitators but also the factors that lead to the enhancement of an environment supportive to quality improvement.

The World State of Quality (WSQ) Project was developed to fill this gap. The Project was conceived to evaluate the macroquality of countries, identifying priorities and areas for improvement. As a result, a national agenda for quality and excellence can be established. The first study in the WSQ project was issued in 2016, covering 28 nations in the European Union (Sampaio et al., 2018; Saraiva et al., 2020) and considering ten different dimensions and 21 indicators. In the following year, the Project was expanded to include nations worldwide. The number of indicators was then reduced to 16, and some of them had to be modified, as required to be adapted to the new global context (Cubo et al., 2019).

The methodology employed by the WSQ Project involves the computation of the Overall World State of Quality (OWSQ) score as a weighted average of the countries' rank positions. The values of the various indicators are not directly used, as they are converted to a normalised scale of ordered ranks. The weights are provided by a panel of international quality specialists. The OWSQ score can be used to compare the different countries and a final rank is generated with the countries with the lowest score representing the top performers. Finally, the countries are grouped by categories according to their rank order. The limits of the categories are set at positions with larger differences between consecutive ranked countries. The five categories created in 2017 were Leader, Follower, Moderate, Lagging, and Beginner, but they were reduced to four in 2018 with the elimination of the category Lagging.

2.2 The use of the Benefit-of-the-Doubt models for constructing composite indicators

A combined analysis of a set of performance indicators is necessary to conduct a multiple-dimensional performance analysis of a set of units for benchmarking purposes. Individual indicators can be aggregated into a single index to create a CI. The Handbook on Constructing Composite Indicators (Nardo et al., 2008) published by the Organisation for Economic Co-operation and Development (OECD) discusses that CIs may accommodate multi-dimensional and complex realities and are often easier to interpret than large sets of individual indicators. However, if improperly developed or misunderstood, they could convey inaccurate signals. Moreover, the selection of indicators and weights could be a source of disagreement, and the aggregation of the various metrics into an overall measure of performance may be challenging, involving creativity and expertise.

Among a large variety of methodological approaches that may be employed for constructing CIs, the Data Envelopment Analysis (DEA) technique has become increasingly popular. DEA is a linear programming technique proposed by Charnes et al. (1978) for evaluating the performance of a group of entities, called "Decision Making Units" (DMUs), that employ multiple inputs to produce multiple outputs. The "Benefit-of-the-Doubt" (BoD) approach was developed by Cherchye et al. (2007) as an application of DEA for building CIs. The BoD technique can overcome the concerns about the need for normalisation and identification of "right" weights, allowing an easy and intuitive interpretation of results. The use of this method based on DEA has the advantage of being data-driven, avoiding an extensive interaction with stakeholders to decide the relative importance of indicators. Furthermore, no normalisation or conversion of units of measurement is necessary for BoD, as the data are used in their original scale. The BoD models have evolved considerably, being adapted to solve different kinds of problems. Various strategies have been applied, for example, to deal with undesirable metrics or to restrict weights within certain bounds.

Weight restrictions may be necessary to prevent the model from using many weight values equal to zero, which may occur in case of a totally free choice of weights. A wide range of weight restriction approaches have been proposed in the DEA literature (Allen et al., 1997; Khalili et al., 2010). The most common type of those restrictions in DEA applications, known as assurance regions type I (ARI), impose limits to ratios between input weights or between output weights.

Once the BoD model is defined as a linear programming problem, it can be solved for each DMU. The optimum solution for the DMU under assessment yields the most favourable weights for that DMU, allowing the estimation of a CI score, ranging from zero to one. After the linear program is computed for all the DMUs, the resulting CI scores enable the ranking of DMUs performance. The best performing DMUs are assigned a CI score equal to one. The hyperplane formed by connecting the efficient DMUs that involves all the data is known as the efficient frontier.

Countries have been used as units of assessment by BoD models in a wide range of applications, such as human development (Rogge, 2018; Van Puyenbroeck and Rogge, 2020), competitiveness (Bowen and Moesen, 2011), social inclusion (Verbunt and Rogge, 2018), environmental performance (Zanella et al, 2013), transportation (Gruetzmacher et al., 2021) and active ageing of population (Amado et al., 2016).

3 DATASET PRESENTATION

The conceptual model of the WSQ Project aims to reflect the quality performance of a country at a broad level including ten dimensions. These dimensions are different in nature, either reflecting the quality-practice approaches of the different countries, or reflecting more intangible concepts, such as "the achievement of sustainable, cohesive and competitive happiness for all that do live or relate with that country" (Saraiva et al., 2020). From each dimension, one or two indicators were chosen from well-known and reliable sources, updated on a regular basis. Some indicators are given in a per capita form to take into account for the country's size. **Erro! A origem da referência não foi encontrada.** displays WSQ dimensions and their respective indicators, with their sources and units of measure (Saraiva et al., 2018).

Table 1 – WSQ Project dimensions and indicators.

DIMENSION	CODE	INDICATOR	UNIT	DATA SOURCE
Organisation	I01	Number of ISO 9001 Certified Organisations	Per 1,000 inhabitants	ISO[1] Survey
Professionals	I02	Number of IAQ Members	Per 1,000 inhabitants	IAQ[2]
Research	I03	Number of Indexed Quality Papers Published	Per 1,000 inhabitants	Scopus, ISI[3]
	I04	Number of Universities in International Research Rankings	Per 1,000 inhabitants	Shangai Ranking
Education	I05	Education Index	score	HDR[4]
Health	I06	Healthy Life Expectancy	age	WHO[5]
	I07	Birth Mortality Rate	Per 1,000 live births	World Bank
Competitiveness	I08	Global Competitiveness Index	score	WEF[6]
	I09	Gross Domestic Product per capita	US dollars	World Bank
Social Cohesion	I10	Gini Index	percentage	World Bank
Sustainability	I11	Number of ISO 14001 Certified Organisations	Per 1,000 inhabitants	ISO
	I12	Ecological Footprint	gha per capita	Footprint Network
Innovation	I13	Global Innovation Index	score	GII[7] Project - WIPO[8]
	I14	Ease of Doing Business Score	score	World Bank
Satisfaction	I15	World Happiness Index	score	SDSN[9]
	I16	Employee Engagement Index	percentage	Gallup

Notes: [1]ISO: International organisation for Standardization; [2]IAQ: International Academy for Quality; [3]ISI: Web of Knowledge, [4]HDR: Human Development Report; [5]WHO: World Health Organisation; [6]WEF: World Economic Forum; [7]GII: Global Inovation Index; [8]WIPO: World Intellectual Property Organisation; [9]SDSN: Sustainable Development Solutions Network.

For the dimension Organisation, the chosen indicator is the number of organisations certified by ISO 9000 (I01), the international set of standards for certification in quality management systems.

The number of members of the International Academy for Quality (IAQ) in each country (I02) is the indicator selected for the dimension Professionals. IAQ is a worldwide non-governmental organisation for professionals dedicated to the development and promotion of quality related methodologies and applications.

In the dimension Research, two metrics were defined. The first one is the number of indexed quality papers published by the country (I03). The sources of that metric are the relevant research databases of ISI (Web of Science) and Scopus, and the number of papers is searched using specific terms in the paper title, abstract and keywords. The second indicator in the dimension Research is the number of the country's universities in the top 500 high-level education institutions listed by the Shanghai ranking (I04).

For the dimension Education, the Education Index (I05), issued annually by the United Nations Development Program was chosen. The Education Index is part of the Human Development Index and represents the average between the expected and the actual number of years of schooling from a country.

Two metrics were selected for the dimension Health: The Health Life Expectancy (I06) and the Birth Mortality Rate (I07). The Healthy Life Expectancy is computed by World Health Organisation and represents the number of years that a person is expected to live in good health conditions aggregated to the country level. The Birth Mortality Rate, issued by the World Bank, is the number of deaths of children before completing one year of life divided per 1,000 live births.

The Global Competitiveness Index (I08) and the Gross Domestic Product per capita (I09) compose the dimension Competitiveness. The Global Competitiveness Index is published by the World Economic Forum and is a CI, including several metrics related to different aspects of competitiveness. The Gross Domestic Product, reported by the World Bank, represents the entire market value of all products and services generated by the economy of a country in a given year.

The dimension Social Cohesion includes the Gini Index (I10), which estimates the deviation from a perfectly equal income within a country. Lower values for this measure indicate a better distribution of income.

In the dimension of Sustainability, two metrics are included: the number of ISO 14001 certified organisations (I11) and the Ecological Footprint (I12). ISO 14001 is the international standard for certification of organisations' environmental management systems and reflects the commitment of organisations to the environmental management of their businesses. The Ecological Footprint expresses the amount of natural resources consumed by the country's population measured in the global hectares (gha) per person, a standardised metric of natural resource consumption. It is computed by the research organisation Global Footprint Network.

The dimension Innovation presents the Global Innovation Index (I13) and the Ease of Doing Business Score (I14). The Global Innovation Index is issued by the World Intellectual Property Organisation (WIPO), a self-funding agency of the United Nations. It is a CI that uses around 80 different indexes, covering several aspects of innovation inputs and outputs. The Ease of Doing Business Score is calculated by the World Bank, and it is also a CI of 41 metrics. The metrics represent regulatory best practices in an economy's country.

Finally, the dimension Satisfaction presents two indicators. The first one is the World Happiness Index (I15), which is annually issued by the Sustainable Development Solutions Network, created by the United Nations to support the implementation of the Sustainable Development Goals. This indicator used data collected by the company Gallup in its World Poll Surveys. The second indicator in the dimension of Satisfaction is the Employee Engagement Index (I16), reported by Gallup, which assesses whether employees are more or less motivated about their jobs depending on how effectively their basic needs are addressed at work.

This work analyses the data published in the World State of Quality report for 2018, which includes 118 countries worldwide, and considers the most recent data available by the mid of the year for each indicator (Saraiva et al., 2018).

The indicators Birth Mortality Index (I07), Gini Index (I10) and Ecological Footprint (I12) present desirable outcomes for lower score values, so they cannot be directly compared to the other measures. They can be converted to positive outcomes, through the replacement of all the data by a large number minus their original values. According to Zanella et al. (2015), the large number should not be exactly the maximum value of the output to ensure the feasibility of the model when weight restrictions are present. Therefore, we used the maximum output of each indicator added to 0.1 for that purpose.

As DEA is sensitive to outliers, it was established that the indicators would range from a minimum value of its average minus three times its standard deviation and a maximum value of its average plus three times its standard deviation. Values beyond this interval were replaced by the range limits. This procedure is recommended by Zanella et al. (2013) to mitigate the impact of extreme values that can bias the performance evaluation. Twelve values were changed in this step, representing 0.6% of the dataset. Those were the only transformations applied to the raw data. The resulting descriptive statistics for the dataset are presented in Table 2.

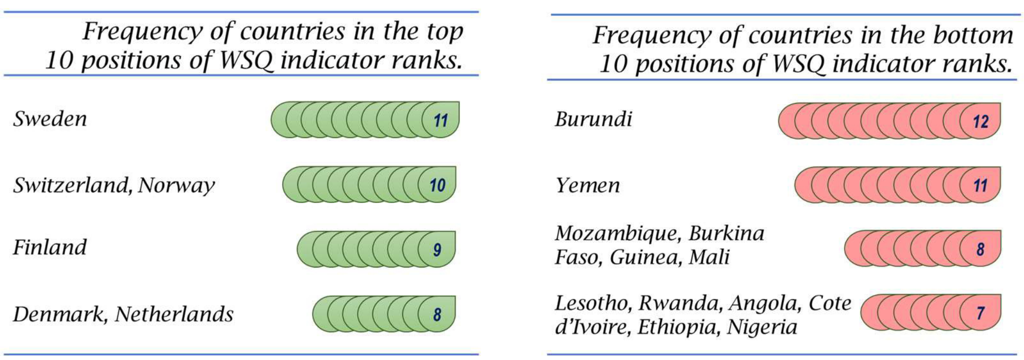
In Table 2, most indicators present high coefficient of variation, reflecting the wide range of performance among the analysed countries.

A further in-depth examination of the dataset indicates a considerable disparity in national performance based on geography. Figure 1 depicts the most often occurring nations at the top and bottom of the 16 indicators' rankings. The highest performing countries are predominantly European (Figure 1a), mainly Scandinavian. Less developed countries, primarily from Africa, are among the lowest performers (Figure 1b).

Table 2 – Descriptive Statistics for all countries in 2018.

CODE	INDICATOR	MEAN	STANDARD DEVIATION	COEFFICIENT OF VARIATION	MINIMUM	MAXIMUM
I01	Number of ISO 9001 Certified Organisations per 1,000 inhabitants	0.24	0.35	1.50	0.00	1.75
I02	Number of members of International Academy for Quality per 1,000 inhabitants	2.99×10^{-5}	7.85×10^{-5}	2.63	0.00	3.80×10^{-4}
I03	Number of Indexed Quality Papers Published per 1,000 inhabitants	0.04	0.06	1.48	0.00	0.22
I04	Number of Universities in International Research Rankings per 1,000 inhabitants	1.18×10^{-4}	2.45×10^{-4}	2.08	0.00	9.00×10^{-4}
I05	Education Index (score)	0.69	0.17	0.24	0.29	0.94
I06	Healthy Life Expectancy (age)	64.70	6.63	0.10	46.60	74.80
I08	Global Competitiveness Index (score)	4.28	0.67	0.16	2.87	5.86
I09	Gross Domestic Product per capita (US dollars)	14192.00	18291.00	1.29	320.00	73272.00
I11	Number of ISO 14001 Certified Organisations per 1,000 inhabitants	0.08	0.12	1.55	0.00	0.57
I13	Global Innovation Index (score)	35.90	12.20	0.34	15.00	68.40
I14	Ease of Doing Business Results (distance to frontier)	65.70	11.50	0.18	31.20	84.10
I15	World Happiness Index (score)	5.46	1.09	0.20	2.90	7.63
I16	Employee Engagement Index (%)	19.00	8.72	0.46	3.30	39.40
I07	Birth Mortality Rate[*1] (per 1,000 live births) – (Complement to 72.5)	53.40	18.20	0.34	0.10	70.70
I10	Gini Index[*1] (%)– (Complement to 64.8)	27.00	8.54	0.32	1.24	48.20
I12	Ecological Footprint[*1] (gha pc) – (Complement to 12.38)	9.22	2.01	0.22	2.90	11.80

[*1] – Undesirable output



(a) Best performers in individual WSQ indicators (b) Worst performers in individual WSQ indicators

Figure 1- Most frequent countries at the top 10 and bottom 10 WSQ individual indicators ranking – 2018.

4 METHODOLOGY

The proposed methodology presents three stages. The first one, in subsection 4.1, details the BoD Model used to estimate the performance of the countries under assessment. This procedure is used as an alternative to the current approach employed in the WSQ Project. The following subsections displays other by-products of the BoD approach. The identification of nations' strengths and shortcomings is described in the subsection 4.2. Finally, the most suitable peers for benchmarking and targets for each indicator are explored in in subsection 4.3.

4.1 BoD model for benchmarking countries

The linear programming model displayed in (1) is known as an output-oriented BoD model and will be employed in this work to construct CIs (Van Puyenbroecket et al., 2021; Zanella et al., 2013; Zanella et al., 2015). In this model, y_{rj} represents the observed value of indicator r for DMU j ($r = 1, \dots, s$ and $j = 1, \dots, n$) and u_r is the weight assigned to indicator r. Note that the set of weights u_r is not known a priori and needs to be determined by optimisation. The letter o denotes the DMU that is being evaluated, and the variable v represents the ratio between the performance of the efficient DMUs and the performance of the DMU under assessment (o). As we are looking for the most favourable weights for o, the objective of the optimisation problem is to minimise v.

$$
\begin{aligned}
& minimise\ v \\
& subject\ to \sum_{r=1}^{s} u_r y_{ro} = 1 \qquad\qquad (1) \\
& \sum_{r=1}^{s} u_r\, y_{rj} - v \leq 0, \quad j = 1, \dots, n \\
& v \geq 0 \\
& u_r \geq 0, \qquad r = 1, \dots, s
\end{aligned}
$$

The optimum solution of the problem (1) yields the weights for DMU o. The value v obtained for the objective function gives the degree to which the benchmark DMUs outperforms DMU o, being greater or equal to one. The value of the composite indicator (CI), that summarises the performance of DMU o, is given by $1/v$ and ranges from zero to one. If the score CI is equal to one, DMU o is itself a benchmark.

Now that we could generate a CI for DMU o, it is worth noting that the same process must be repeated for all the DMUs. Therefore, the linear programming problem must be solved n times for all the DMUs under consideration. At the end of this process, all DMUs are assigned a CI score, all ranging from zero to one, so they can be compared to each other. In each case, a set of most-favourable weights will also be produced, indicating the trade-offs between the various indicators for the country being evaluated.

In linear programming theory, every linear programming problem has an alternative problem derived from it, which provides equivalent solutions, but uses different decision variables. Primal refers to the initial linear programming problem, whereas the derived linear programming problem is known as dual. The dual formulation of model (1) is presented as model (2). The primal formulation in BoD is often referred as weights' formulation and the dual is known as envelopment formulation. The envelopment formulation in BoD models is often used to identify peers and targets of the DMUs under consideration for benchmarking purposes (Zanella et al., 2013; Pereira et al., 2021; Henriques et al., 2020; Oliveira et al., 2020).

In model (2), the decision variables λ_j mean the degree of similarity between the DMU under consideration and its peers. The factor θ by which all outputs of the DMU under evaluation can be proportionally raised to meet the target values corresponds to the objective function of model (2). The value of CI can be also obtained from model (2) as $1/\theta$, and it is equivalent to the CI given by model (1).

$$
\begin{aligned}
& maximise\ \theta \\
& subject\ to\ \ \theta y_{ro} - \sum_{j=1}^{n} \lambda_j y_{rj} \leq 0, \qquad r = 1, \dots, s \qquad (2) \\
& \sum_{j=1}^{n} \lambda_j \leq 1, \\
& \lambda_j \geq 0, \qquad j = 1, \dots, n \\
& \theta\ \ is\ free
\end{aligned}
$$

In the model displayed in (1), the weights can be chosen freely during the optimisation process. That could lead to the generation of weights equal to zero for some of the indicators, and, therefore, the importance of some dimensions can be neglected. To overcome this limitation, constraints must be added to the model to limit the range of the weights produced for each indicator or for each dimension being assessed.

In this work, two alternatives for weight restrictions are applied. In both alternatives, an "artificial" country is considered in the proposed strategy, presenting the average performance among all countries. Therefore, the indicators' values of this country are equivalent to the average of all countries' indicators ($\overline{y}_r$). In the first option presented in (3), we impose that the weights of the "artificial" country match the original weights provided by the WSQ Project (w_r), presented in Table

3. This option is used to illustrate the difference between the BoD approach and the weighted average method using rank positions employed in the WSQ Project.

$$\frac{u_r \overline{y}_r}{\sum_{r=1}^{s} u_r \, \overline{y}_r} = w_r, \qquad r = 1 \dots s \qquad (3)$$

Table 3 - Fixed weights used in the WSQ Project.

Indicator	I01	I02	I03	I04	I05	I06	I07	I08	I09	I10	I11	I12	I13	I14	I15	I16
Weight (w_r) (%)	6.00	4.80	6.24	5.94	5.93	6.40	6.24	6.63	6.41	5.67	6.09	6.20	6.61	6.59	7.69	6.54

A more flexible approach using ARI restrictions is proposed in (4), using a lower bound ϕ_r for all the indicators. Establishing a lower threshold ϕ_r can avoid the generation of zero weights and gives more freedom for the model to determine optimum solutions.

$$\frac{u_r \overline{y}_r}{\sum_{r=1}^{s} u_r \, \overline{y}_r} \geq \phi_r, \qquad r = 1 \dots s \qquad (4)$$

The alternative approaches for weight restrictions using (3) or (4) can be added to model (1) to generate different perspectives for countries performance assessment.

4.2 Identification of strengths and weaknesses

Using the weights assigned by the BoD model, it is possible to analyse which indicators are given higher importance in the performance of the countries under assessment. The CI score can be split into indicator contributions representing virtual weights. The virtual weights are obtained as the product of each indicator's value (y_{ro}) by its weight (u_r) given by the model results, and the virtual weights of a given country sum up to one as shown in the first restriction of model (1).

For the identification of strengths and weaknesses, the outcomes generated from the BoD model with constraints (4) are employed. Since the lower bound for the weights is ϕ_r on average, we assume that the model provides virtual weights higher than ϕ_r aiming to favour the country under assessment for the indicators with best performance. Those represent the country's strengths, while the indicators with low virtual weights represent the country's weaknesses. In the case of the OWSQ original method, considering that the score is calculated by multiplying the fixed weights by the indicator's ranking position, the final score represents a weighted average of the country's ranking positions for each indicator. Therefore, in this situation, it is not possible to compute the contribution of each indicator to the country's overall score.

4.3 Identification of peers and targets

The BoD models can be used to identify the country's peers for benchmarking purposes. The peers are other countries whose results a country should consider to enhance its performance. In order to look for more comparable peers and achievable targets, we choose to run the BoD models for country clusters and identify efficient frontiers within homogeneous groups. The clusters were made up of countries with comparable geographical locations. The use of geography as an exogenous variable here is justified by the large variations in countries' performances associated with this factor, as previously discussed.

The λ_j values obtained from the BoD model (2) within each cluster indicate the measure of similarity between a country and its peers. The values of λ_j that are different from zero identify the country's peers. The targets for each indicator are taken from the indicator values of the peers, that outperforms the country under assessment.

5 RESULTS AND DISCUSSION

This section is divided in three parts. Subsection 5.1 presents the results of countries' ranks and the proposed categorisation, discussing the differences between OWSQ and BoD approaches. Subsection 5.2 discusses nations' strengths and weaknesses that can be extracted from the BoD model, with illustrative examples of two countries. The same two countries are used as examples to discuss the identification of peers and targets in Subsection 5.3.

5.1 The overall ranking of countries

Overall CIs are generated for all the countries, using the two different approaches for weight restrictions in the BoD model. The first one imposes that an "average" DMU uses the same weights employed to generate the original Overall World State of Quality (OWSQ) score. The second one is more flexible and forces the "average" DMU to present weights that are higher or equal to ϕ_r for each indicator, according to expression (4). The value of ϕ_r was chosen as 4% (0.04) for all the indicators in this work. This value was selected, because it allows that all 16 indicators present significant contributions of at least 64% and still there is some flexibility (36%) for searching for the most favourable weights in each country.

The CIs and rank positions computed from the BoD models and a comparison with the Overall World State of Quality results (OWSQ) are shown in Table A.1 in Appendix A. In this table, the countries are also presented with their three-digit codes, according to the international standard ISO 3166. We categorise the nations by dividing the BoD flexible-weight rank into quartiles, with 29 elements in the first quartile, 30 in the second, 29 in the third and 30 countries in the fourth quartile. As a result, the first quartile, covering the top 25% higher positions in the rank, represents the category Leader. Countries in the following quartiles are designated as Followers, Moderates, and Beginners, in this order. The use of quartiles to categorize the countries seems to be a more viable approach in this case, because the larger differences between subsequent nations in the BoD CI ranking, which is the original WSQ categorization criteria, are concentrated within the top 10 rank positions.

A comparison between the OWSQ rank and BoD ranks is performed using Spearman's rank correlation technique. When the OWSQ rank and BoD rank using the same weights are compared, the Spearman's rho coefficient is 0.984, showing a strong correlation between the two ranks. The resulting p-value is less than 0.001 indicating that the correlation is statistically significant. The results of applying OWSQ weights in the BoD model show that both ranks are comparable, although no normalisation procedure was used. In this context, information loss is minimised by utilising actual data rather than ranking positions. A similar analysis between the OWSQ rank and BoD flexible-weight approach yields a Spearman's rho coefficient of 0.840 and a p-value of less than 0.001. The correlation is still high and statistically significant in this scenario, even though the weights are allowed to vary aiming to highlight countries' best outcomes.

The BoD model with ARI weight restrictions at a minimum of 4% is used for the following analyses. A scatter plot showing the correlation between the OWSQ rank and the BoD rank positions is presented in Figure 2. The shaded area around the trend line shows the 95% confidence interval for the correlation. The categories assigned to the OWSQ classification are colour-coded in the graph and the categories according to the BoD ranking positions are shown in the graph area limited by the dotted horizontal lines.

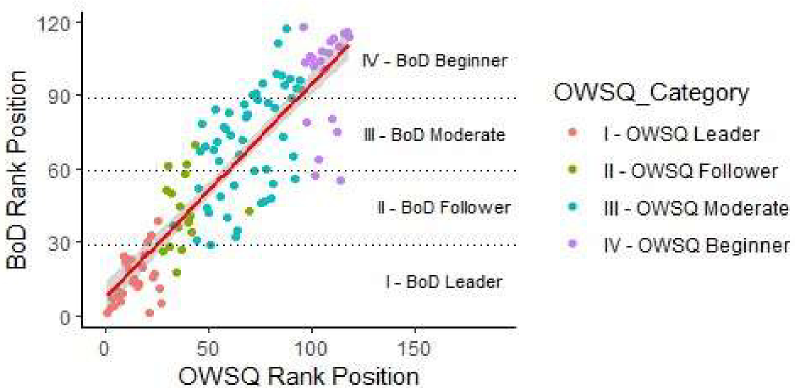
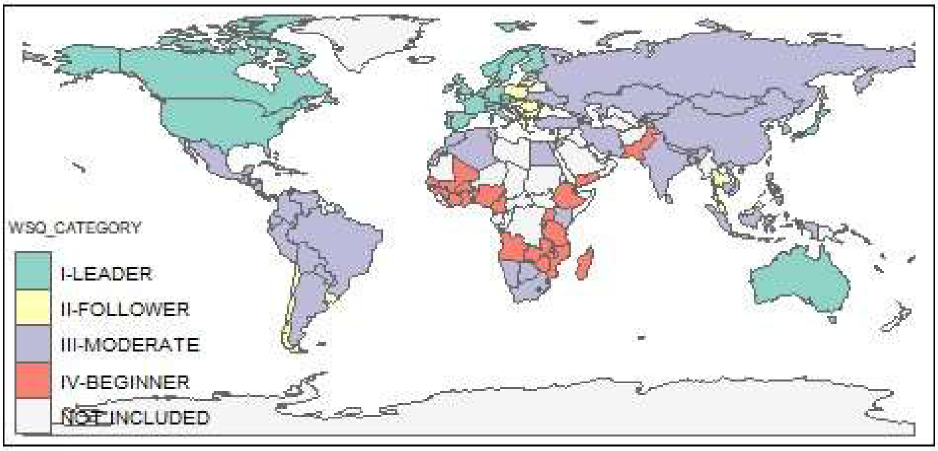
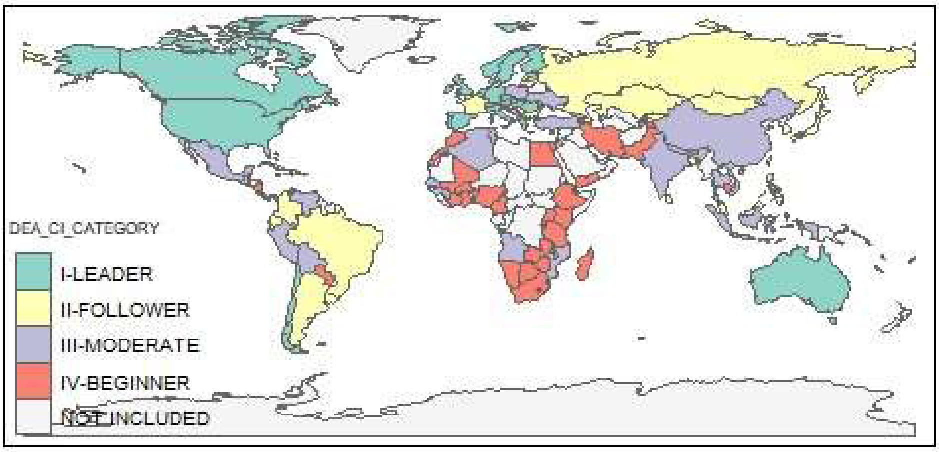


Figure 2 - Scatter plot of BoD and OWSQ rank positions.

In Figure 2, some variation between the original OWSQ classification and the BoD categories can be noticed. The similarities and differences between both classification methods can be better illustrated in the world maps in Figure 3.



(a) OWSQ classification - 2018



(b) BoD Classification - 2018

Figure 3 - Classification of countries using two different approaches.

The OWSQ categories are unbalanced in terms of the number of elements, with the category Moderate presenting 52 countries, while Followers exhibit only 16 members. Leaders and Beginners present 27 and 23 countries, respectively. On the other hand, by dividing the nations into quartiles, this work decides to keep the number of members similar in each group. This fact explains part of the discrepancies that can be observed in Figure 3. For example, in Figure 3, Brazil, Argentina, and Russia shift from the OWSQ category of Moderate to the BoD classification of Follower. However, as seen in Table A.1, the rank positions of those nations did not alter significantly. Brazil is ranked 56th, Russia is 45th, and Argentina is 49th in the OWSQ approach. The BoD rank positions for these countries are 49th, 52nd, and 44th, respectively. That means that the changes in the number of elements in each category play a major role at the differences between the two categorisation methods.

A detailed visualization in the Africa map presented in Figure 4 exemplifies some discrepancies between the two classification strategies. This figure shows a set of Beginners in the OWSQ approach that present better classifications in the BoD technique. Two of them, Burundi and Lesotho, change to Follower and Angola, Guinea, Senegal and Mozambique shift to Moderate. The African countries typically present better outcomes for the Ecological Footprint (I12), notably due to the lower consumption of natural resources by their least developed economies. However, this specific group of countries additionally benefit from their performance in the Employee Engagement Index (I16), even if compared with the other nations from Africa. This effect explains the shift in category, a shift that is not evident in the OWSQ approach.

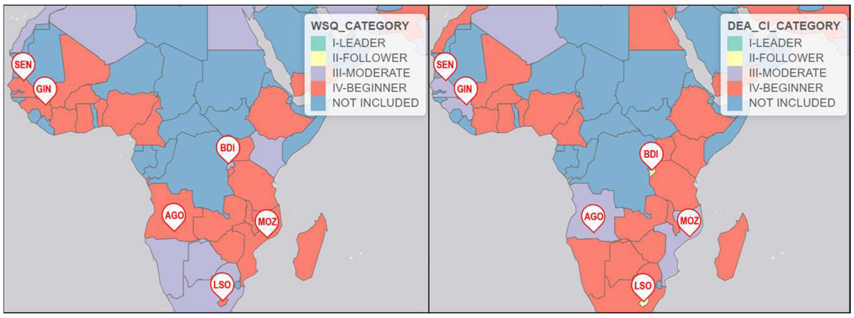


Figure 4 - Differences in OWSQ and BoD classification approaches - Examples in African countries.

Other discrepancies that can be explained by the use of different methodologies appear in a closer look in European countries. Figure 5 depicts the main changes between Leaders and Followers in Europe comparing OWSQ and BoD techniques. France switched from being a Leader in the OWSQ technique to being a Follower in the BoD method. In contrast, Hungary, Greece, and Romania switched from being Followers in the OWSQ method to being Leaders in the BoD method.

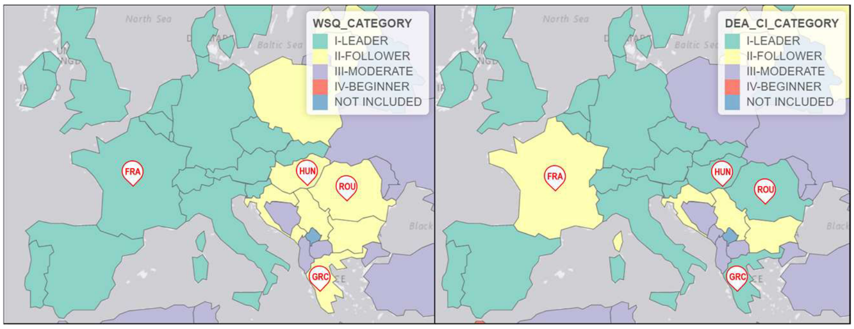


Figure 5 - Differences in OWSQ and BoD classification approaches - Examples in European countries.

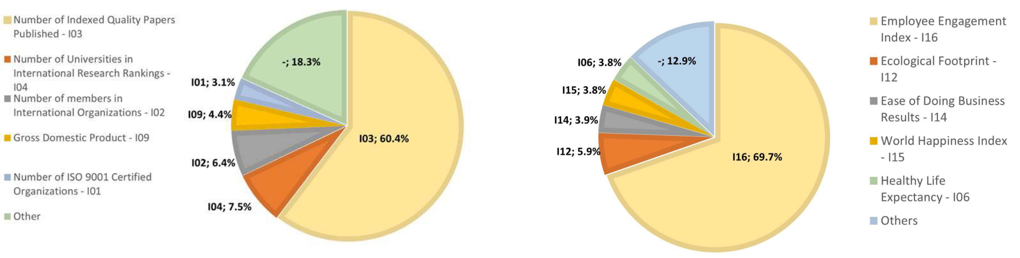
Considering a direct comparison between France and Hungary, Table 4 shows that France dropped from 22nd in OWSQ rank to 33rd in BoD, whereas Hungary rose from 34th in OWSQ to 18th in BoD approach. Examining the data with a higher level of detail, in Table 4, one can see the values of all observations and ranking order for all indicators in both countries. France outperforms Hungary in 11 out of the 16 indicators, but the order of magnitude of the value differences is lost when the values are converted in rank positions. In some cases, large discrepancies in ranking such as in Life Expectancy (I06), where France ranks 4 and Hungary ranks 50, account for less than 10% of the observation's difference. Another example is the Birth Mortality Rate, where a difference of 12 rank positions (20th for France, 32nd for Hungary) represents only 1.9% in the values. Looking at the indicators where Hungary excels, the situation of the Number of members of International Academy for Quality (I02) has the opposite impact. Hungary's performance is 20 times that of France, however the gap in rankings is just 21 places: Hungary is fourth and France is 25th. Therefore, the use of ranking positions to compose the OWSQ score plays a significant role in the differences between OWSQ and BoD techniques.

Table 4 - Comparison between performances of France and Hungary.

Code	Indicator	Observed value - 2018		Indicator Ranked Order	
		FRA	HUN	FRA	HUN
I01	Number of ISO 9001 Certified Organizations per 1,000 inhabitants	0.35	0.67	31	14
I02	Number of members of International Academy for Quality per 1,000 inhabitants	1.49×10^{-5}	3.07×10^{-4}	25	4
I03	Number of Indexed Quality Papers Published per 1,000 inhabitants	0.06	0.04	33	35
I04	Number of Universities in International Research Rankings per 1,000 inhabitants	2.98×10^{-4}	0	19	41
I05	Education Index (score)	0.84	0.82	25	32
I06	Healthy Life Expectancy (age)	73.40	66.80	4	50
I08	Global Competitiveness Index (score)	5.18	4.33	17	49
I09	Gross Domestic Product per capita (US dollars)	38476.66	14224.85	17	37
I11	Number of ISO 14001 Certified Organizations per 1,000 inhabitants	0.10	0.23	32	15
I13	Global Innovation Index (score)	54.40	44.90	14	29
I14	Ease of Doing Business Results (distance to frontier)	76.13	72.39	25	40
I15	World Happiness Index (score)	6.49	5.62	20	54
I16	Employee Engagement Index (%)	6.20	10.30	113	102
I07	Birth Mortality Rate (per 1,000 live births) - (Complement to 72.5)	69.30	68.10	20	32
I10	Gini Index (%) - (Complement to 64.8)	32.10	34.40	39	20
I12	Ecological Footprint (gha pc) - (Complement to 12.38)	7.69	8.78	93	77

5.2 Identification of strengths and weaknesses

We choose two nations with different performances as examples to analyse the different contributions of each indicator to their performance. The Netherlands' virtual weights are presented in Figure 6a, and Figure 6b represents the contributions of the indicators to Mozambique's performance.



(a) Virtual weights - The Netherlands (b) Virtual weights - Mozambique

Figure 6: Contributions of the indicators for Mozambique and The Netherlands.

The most relevant strength of the Netherlands (Figure 6a) is the number of quality papers published (I03), which contributes to more than 60% of its performance, followed by the number of universities in international ranks (I04). These two metrics in the dimension Education account for 67.9% of the Netherlands' performance, demonstrating the country's attention in this field. Other strengths above the 4% threshold are the number of members in IAQ (I02) and the Gross Domestic Product (I09). Its main weaknesses are the Ecological Footprint (I12) and the Employee Engagement Index (I16) which contribute with less than 1% for the final score.

The example of Mozambique in Figure 6b shows that the most significant strength of this country is the Employee Engagement Index (I16), followed by the Ecological Footprint (I12), as previously discussed in Subsection 5.1. All the remaining virtual weights of Mozambique are below 4% indicating its poor performance compared to the efficiency frontier built by the BoD model. The Employee Engagement Index (I16) and the World Happiness Index (I15), both from the dimension Satisfaction, contribute for 73.5 percent of Mozambique's overall score, reflecting the country's achievements in this area.

5.3 Identification of peers and targets

Continuing with the Netherlands and Mozambique as examples, two separate clusters are chosen based on the geographic areas in which the nations are located, as defined by the United Nations Statistics Division (United Nations 2021). Two different BoD models will be computed in this analysis, one for each cluster. The clusters are used to create more uniform groups of countries for the identification of peers and targets so that the BoD model can be resolved for this smaller group.

In the case of the Netherlands, the model was computed using all the countries from Europe included in the WSQ Project. The CI scores and ranks for the countries under assessment in this cluster are presented in Table B.1 in Appendix B.

A BoD model was also solved considering the set of countries from the region of Sub-Saharan Africa, which includes Mozambique. Table B.2 in Appendix B displays the countries in this cluster, their CI scores and rank positions.

Table 5 presents the peers of the Netherlands and Mozambique, determined by the values of λ_j obtained from the results of optimisation model (2).

Table 5 - Peers and Targets Identification - Examples of the Netherlands and Mozambique for 2018.

Code	Indicator	Netherlands 2018 results	Peers: Finland ($\lambda_j = 0.28$)	Peers: Switzerland ($\lambda_j = 0.82$)	Mozambique 2018 results	Peer: South Africa ($\lambda_j = 1$)
I01	Number of ISO 9001 Certified Organizations per 1,000 inhabitants	0.603	0.470	1.324	0.002	0.084
I02	Number of members of International Academy for Quality per 1,000 inhabitants	1.75×10^{-4}	3.79×10^{-4}	3.79×10^{-4}	0	3.53×10^{-5}
I03	Number of Indexed Quality Papers Published per 1,000 inhabitants	0.20	0.16	0.22	0	0.02
I04	Number of Universities in International Research Rankings per 1,000 inhabitants	7.00×10^{-4}	9.01×10^{-4}	9.01×10^{-4}	0	8.82×10^{-5}
I05	Education Index (score)	0.91	0.91	0.90	0.39	0.71
I06	Healthy Life Expectancy (age)	72.10	71.70	73.50	52.20	55.70
I07	Birth Mortality Rate (per 1,000 live births)	3.20	1.90	3.60	53.10	34.20
I08	Global Competitiveness Index (score)	5.66	5.49	5.86	2.89	4.32
I09	Gross Domestic Product per capita (US dollars)	48,223.16	45,703.33	73,271.97	415.72	6,160.73
I10	Gini Index (%)	29.30	27.10	32.50	54.00	63.00
I11	Number of ISO 14001 Certified Organizations per 1,000 inhabitants	0.16	0.26	0.37	0	0.02
I12	Ecological Footprint (gha pc)	5.92	6.09	4.85	0.87	3.42
I13	Global Innovation Index (score)	63.30	59.60	68.40	23.10	35.10
I14	Ease of Doing Business Results (distance to frontier)	76.03	80.37	75.92	54.00	64.89
I15	World Happiness Index (score)	7.44	7.63	7.49	4.42	4.72
I16	Employee Engagement Index (%)	12.20	12.20	13.20	28.00	15.30

The Netherlands' peers are Finland and Switzerland, which are considered efficient countries in the set selected. The Netherlands needs to look for the best practices of Finland and Switzerland, as those countries present a similar performance in macroquality compared to the Netherlands. The values of λ_j give the degree of similarity of the Netherlands and its peers. The targets for each indicator are determined looking at the results of the country's peers. For example, the Netherlands needs to nearly double its results for the number of members of International Academy for Quality (I02) to reach the values of Finland and Switzerland (3.79×10^{-4} members per 1,000 inhabitants), which represents the Netherlands' target for this indicator. In this case, both peers present better performance and may offer valuable insights on reaching the target. Notice that, even though I02 is considered a strength for the Netherlands in a worldwide context, a comparison within a more homogeneous group reveals other peers with better performance. In the case of the Global Innovation Index (I13), for instance, the results of the Netherlands are better than Finland's. Therefore, the Netherlands should look for Switzerland's practices to investigate potential improvements for this indicator. In this case, the Netherlands' target is 68.4, which is Switzerland's results.

The BoD model for the Sub-Saharan countries, only one country is identified as Mozambique's peer: South Africa. In this context, compared with Mozambique, the value of λ_j for South Africa equals one. Mozambique outperforms South Africa only in Employee Engagement Index (I16), Gini Index (I10) and Ecological Footprint (I12), so for all the other metrics, South Africa's results represent Mozambique's targets, and Mozambique may learn from South Africa's practices.

This kind of analysis represents one of the main advantages of the BoD technique compared to other strategies, since it may successfully provide direction on how the performance of a DMU can be enhanced by looking at best practices from their counterparts.

6 CONCLUSIONS

This work discusses an approach based on the Benefit-of-the-Doubt (BoD) technique to assess countries regarding their performance in quality, as an alternative to the method employed by the World State of Quality (WSQ) Project. A composite indicator (CI) resulting from the aggregation of 16 metrics is computed and the countries are ranked and categorised based on that. The similarities and discrepancies between both methods are discussed using the results of some countries as example. The results show that the BoD approach and the original method used by the WSQ Project present comparable results. Therefore, the proposed method can be alternatively employed for the WSQ Project presenting the advantage of being data-driven, so it is not necessary to judge the relative importance of the various metrics. In that sense, no country can complain about the resulting weights because they are obtained in the most favourable manner. Other than that, the data can be employed in their original form, requiring no transformations that can affect the analysis.

Another benefit of the BoD strategy is the identification of a country's strengths and weaknesses, which is discussed using illustrative examples. Furthermore, this strategy allows for the establishment of appropriate targets to steer the development of a country's performance as well as the identification of peers to serve as benchmarks. Countries from diverse geographical areas were presented as examples.

Future studies might look at using different measures to emphasize the many facets of a country's quality performance. Other research opportunities would be to look at the evolution of quality performance through time and to explore different approaches for countries' categorisation, such as a segmentation by continents.

ACKNOWLEDGEMENTS

The authors acknowledge the financial support provided by FCT- Fundação para a Ciência e a Tecnologia (Portuguese National Funding Agency for Science, Research and Technology) through PhD research grants 2021.05244.BD and SFRH/BD/131285/2017.

This work has been supported by FCT – Fundação para a Ciência e Tecnologia within the R&D Units Project Scope: UIDB/00319/2020.

REFERENCES

Ahmed, S. M., Aoieong, R. T., Tang, S. L., and Zheng, D. X. M. (2005), “A comparison of quality management systems in the construction industries of Hong Kong and the USA”, International Journal of Quality and Reliability Management, Vol. 22 No. 2, pp. 149–161.

Aiginger, K. (1998), “A Framework for Evaluating the Dynamic Competitiveness of Countries”, Structural Change and Economic Dynamics, Vol. 9 No. 2, pp. 159–188.

Allen, R., Athanassopoulos, A., Dyson, R. G., and Thanassoulis, E. (1997), “Weights restrictions and value judgements in Data Envelopment Analysis: Evolution, development and future directions”, Annals of Operations Research, Vol. 73, Chapter 2, pp. 13–34.

Amado, C. A. F., São José, J. M. S., and Santos, S. P. (2016), “Measuring active ageing: A Data Envelopment Analysis approach”, European Journal of Operational Research, Vol. 255 No 1, pp. 207–223.

APQC, and ASQ (2013), “The ASQ Global State of Quality – Research – Discoveries 2013”, United States of America: American Society for Quality.

——— (2016), “The ASQ Global State of Quality – Research 2 – Discoveries 201.”, United States of America: American Society for Quality.

Bello-Pintado, A., and Merino-Díaz-de-Cerio, J. (2013), “Determinants of the use of quality management practices in Latin America: The case of Argentina and Uruguay”, Total Quality Management and Business Excellence, Vol. 24 No 1-2, pp. 31–47.

Bou-Llusar, J. C., Escrig-Tena, A. B., Roca-Puig, V., and Beltrán-Martín, I. (2009), “An empirical assessment of the EFQM Excellence Model: Evaluation as a TQM framework relative to the MBNQA Model”, Journal of Operations Management, Vol. 27 No 1, pp. 1–22.

Bowen, H. P., and Moesen, W. (2011), “Composite competitiveness indicators with endogenous versus predetermined weights: an application to the World Economic Forum’s Global Competitiveness Index”, Competitiveness Review: An International Business Journal.

Charnes, A., Cooper, W. W., and Rhodes, E. (1978), “Measuring the efficiency of decision making units”, European Journal of Operational Research, Vol. 2 No 6, pp. 429–444.

Chen, Y. Y., Wu, L., and Zhai, Q. G. (2019), “Does ISO 9000 certification benefit service firms?”, Sustainability (Switzerland), Vol. 11 No 21, pp. 1–22.

Cherchye, L., Moesen, W., Rogge, N., and Puyenbroeck, T. V. (2007), “An introduction to ‘Benefit of the Doubt’ composite indicators”, Social Indicators Research, Vol. 82 No 1, pp. 111–145.

Clegg, B., Roya G., and Mine O. (2013), "Quality management and performance: A comparison between the UK and Turkey", Production Planning and Control, Vol. 24 No 12, pp. 1015–1031.

Cubo, C., Saraiva, P., Sampaio, P., and Reis, M. (2019), "2017 World State of Quality: First Worldwide Results", Total Quality Management and Business Excellence, Vol. 32 No 304, pp. 1–10.

Dahlgaard, J. J., Kristensen, K., Kanji, G. K., Juhl, H. J., and Sohal, A. S. (1998), "Quality management practices: A comparative study between East and West", International Journal of Quality and Reliability Management, Vol. 15 No 8, pp. 812–826.

Dutta, S., Lanvin, B., and Wunsch-Vincent, S. (2020), Global Innovation Index 2020. Johnson Cornell University.

Farris, F. A. (2010), "The Gini index and measures of inequality", The American Mathematical Monthly, Vol. 117 No 10, pp. 851–864.

Gruetzmacher, S. B., Vaz, C. B., and Ferreira, A. P. (2021), "Performance Assessment of the Transport Sustainability in the European Union", Communications in Computer and Information Science, Vol. 1359, pp. 106–119.

Helliwell, J. F, Layard, R., Sachs, J. D., and De Neve, J. (2020), "World Happiness Report 2020.", New York: Sustainable Development Solutions Network.

Henriques, A. A., Camanho, A. S., Amorim, P., and Silva, J. G. (2020), "Performance benchmarking using composite indicators to support regulation of the Portuguese wastewater sector", Utilities Policy, Vol. 66: 101082.

Ismail, M. Y., M. Baradie, El, and Hashmi, M. S. J. (1998), "Quality Management in the Manufacturing Industry: Practice Vs Performance", Computers and Industrial Engineering, Vol. 35 No 3-4, pp. 519–522.

Jaeger, M., Adair, D., and Al-Qudah, S. (2013), "MBNQA criteria used in the GCC countries", The TQM Journal, Vol. 25 No 2, pp. 110–123.

Khalili, M., Camanho, A. S., Portela, M. C. A. S., and Alirezaee, M. R. (2010), "The measurement of relative efficiency using Data Envelopment Analysis with assurance regions that link inputs and outputs", European Journal of Operational Research, Vol. 203 No 3, pp. 761–770.

Lau, R. S. M., Zhao, X., and Xiao M. (2004), "Assessing quality management in China with MBNQA criteria", International Journal of Quality & Reliability Management.

Nardo, M., Saltelli, A., Tarantola S., Hoffmann A., and Giovannini E. (2008), Handbook on Constructing Composite Indicators: Methodology and User Guide. OECD; European Union; Joint Research Centre - European Commission.

Oliveira, R., Zanella A., and Camanho A. S (2020), “A temporal progressive analysis of the social performance of mining firms based on a Malmquist index estimated with a Benefit-of-the-Doubt directional model”, Journal of Cleaner Production, Vol. 267: 121807.

Pereira, M. A., Camanho, A. S., Figueira, J. R., and Marques R. C. (2021). “Incorporating preference information in a range directional composite indicator: The case of Portuguese public hospitals”, European Journal of Operational Research, Vol. 294 No 2, pp. 633–650.

Qazi, T. F., Niazi A. A. K., Asghar W., and Basit A. (2021), “Ease of Doing Business: analysis of trade facilitations of one hundred twenty-seven countries of the world”, Journal of Accounting and Finance in Emerging Economies, Vol. 7 No 1, pp. 65–75.

Rogge, N. (2018), “Composite indicators as generalized benefit-of-the-doubt weighted averages”, European Journal of Operational Research, Vol 267 No 1, pp. 381–392.

Rogge, N., and Archer G. (2021), “Measuring and analyzing country change in establishing ease of doing business using a revised version of World Bank’s Ease of Doing Business Index”, European Journal of Operational Research, Vol. 290. pp. 373–385.

Sampaio, P., Saraiva P., Cubo C. and Reis, M. (2018). “From another quality dimension”, Quality Progress, Vol. 51 No 12, pp. 50–57.

Saraiva, P., Sampaio P., Cubo C., and Reis M. (2020), “Macroquality Measurement: World State of Quality and European Quality Scoreboard approaches and results”, Total Quality Management and Business Excellence, Vol. 31, No 9-10, pp. 1060–1076.

Saraiva, P., Sampaio, P., Cubo, C., Reis, M., & d’Orey, J. (2018). 2018 World State of Quality (Final Report). University of Minho and University of Coimbra, Portugal. Available at http://wsq.dps.uminho.pt/.

United Nations (2021). “UNSD - Methodology. Standard country or area codes for statistical use (M49). UN Department of Economic and Social Affairs. Statistics Division. Standard Country or Area Codes for Statistical Use (M49). UN Department of Economic; Social Affairs. Statistics Division.

Van Iwaarden, J., Van Der Wiele, T., Dale, B., Williams, R., and Bertsch, B. (2008), “The Six Sigma improvement approach: a transnational comparison”, International Journal of Production Research, Vol. 46 No 23, pp. 6739–6758.

Van Puyenbroeck, T., Montalto V., and Saisana M. (2021), "Benchmarking culture in Europe: A Data Envelopment Analysis approach to identify city-specific strengths", European Journal of Operational Research, Vol. 288 No 2, pp. 584–597.

Van Puyenbroeck, T., and Rogge N. (2020), "Comparing regional human development using global frontier difference indices", Socio-Economic Planning Sciences, Vol. 70, pp. 1–8.

Verbunt, P., and Rogge N. (2018), "Geometric composite indicators with compromise Benefit-of-the-Doubt weights", European Journal of Operational Research, Vol. 264 No 1, pp. 388–401.

Zanella, A., Camanho A. S., and Dias T. G. (2013), "Benchmarking countries' environmental performance", Journal of the Operational Research Society, Vol. 64 No 3, pp. 426–438.

Zanella, A., Camanho A. S., and Dias, T. G. (2015), "Undesirable outputs and weighting schemes in composite indicators based on data envelopment analysis", European Journal of Operational Research, Vol. 245 No 2, pp 517–530.

APPENDICES

Appendix A - Comparison between BoD and OWSQ techniques

Table A.1 - Composite Indicator - Comparison BoD and OWSQ techniques.

		OWSQ Approach			BoD using OWSQ weights		BoD using flexible weights ≥ 4%		
Country	Code	Score	Rank Position	Category	Composite Indicator	Rank Position	Composite Indicator	Rank Position	Category
Albania	ALB	52.82	52	Moderate	0.31	58	0.40	68	Moderate
Algeria	DZA	71.39	82	Moderate	0.26	88	0.36	85	Moderate
Angola	AGO	85.08	103	Beginner	0.23	105	0.42	64	Moderate
Argentina	ARG	50.56	49	Moderate	0.38	40	0.49	44	Follower
Armenia	ARM	60.02	65	Moderate	0.30	67	0.41	66	Moderate
Australia	AUS	23.52	7	Leader	0.75	4	0.88	6	Leader
Austria	AUT	26.89	12	Leader	0.62	16	0.66	23	Leader
Azerbaijan	AZE	63.50	72	Moderate	0.27	80	0.44	59	Follower
Bangladesh	BGD	75.33	92	Moderate	0.27	81	0.44	56	Follower
Belgium	BEL	30.66	18	Leader	0.63	14	0.70	20	Leader
Benin	BEN	93.02	117	Beginner	0.19	118	0.31	116	Beginner
Bhutan	BTN	78.69	96	Beginner	0.23	103	0.29	118	Beginner
Bolivia	BOL	75.10	91	Moderate	0.26	89	0.41	65	Moderate
Bosnia-Herzegovina	BIH	56.62	58	Moderate	0.33	49	0.39	76	Moderate
Botswana	BWA	72.74	84	Moderate	0.25	96	0.33	111	Beginner
Brazil	BRA	55.68	56	Moderate	0.32	52	0.47	49	Follower
Bulgaria	BGR	44.19	40	Follower	0.43	33	0.51	38	Follower
Burkina Faso	BFA	92.76	116	Beginner	0.20	117	0.32	115	Beginner
Burundi	BDI	88.90	114	Beginner	0.23	101	0.45	55	Follower
Cambodia	KHM	75.64	93	Moderate	0.26	85	0.36	93	Beginner
Cameroon	CMR	88.54	113	Beginner	0.21	110	0.33	110	Beginner
Canada	CAN	26.89	11	Leader	0.58	19	0.71	19	Leader
Chile	CHL	40.24	31	Follower	0.40	37	0.56	28	Leader
China	CHN	53.03	54	Moderate	0.34	44	0.40	71	Moderate
Colombia	COL	48.78	44	Moderate	0.35	42	0.54	31	Follower
Costa Rica	CRI	44.22	41	Follower	0.34	45	0.50	41	Follower
Cote d'Ivoire	CIV	87.18	109	Beginner	0.20	115	0.32	112	Beginner
Croatia	HRV	41.15	35	Follower	0.46	30	0.51	36	Follower
Cyprus	CYP	37.51	29	Follower	0.41	35	0.46	51	Follower
Czech Republic	CZE	27.82	13	Leader	0.54	22	0.74	14	Leader
Denmark	DNK	21.95	5	Leader	0.68	8	0.82	8	Leader
Dominican Republic	DOM	67.22	77	Moderate	0.29	72	0.47	47	Follower
Ecuador	ECU	58.46	62	Moderate	0.30	62	0.46	53	Follower
Egypt	EGY	72.13	83	Moderate	0.27	82	0.35	99	Beginner
El Salvador	SLV	63.34	70	Moderate	0.30	63	0.49	43	Follower
Estonia	EST	29.03	16	Leader	0.62	15	0.75	13	Leader
Ethiopia	ETH	89.96	115	Beginner	0.22	108	0.33	109	Beginner
Finland	FIN	20.83	4	Leader	0.84	3	0.94	4	Leader
France	FRA	32.89	22	Leader	0.48	28	0.52	33	Follower
Georgia	GEO	53.05	55	Moderate	0.31	55	0.43	63	Moderate
Germany	DEU	24.17	9	Leader	0.61	18	0.65	24	Leader
Ghana	GHA	80.46	99	Beginner	0.24	97	0.33	106	Beginner
Greece	GRC	42.23	37	Follower	0.51	26	0.56	27	Leader
Guatemala	GTM	67.62	78	Moderate	0.28	76	0.44	60	Moderate
Guinea	GIN	87.63	112	Beginner	0.22	109	0.39	75	Moderate
Honduras	HND	69.99	81	Moderate	0.28	75	0.46	54	Follower
Hungary	HUN	40.57	34	Follower	0.56	21	0.71	18	Leader
India	IND	68.24	79	Moderate	0.26	87	0.36	87	Moderate
Indonesia	IDN	65.07	74	Moderate	0.28	79	0.36	88	Moderate
Iran	IRN	63.60	73	Moderate	0.30	66	0.36	91	Beginner
Ireland	IRL	23.73	8	Leader	0.68	7	0.81	9	Leader
Israel	ISR	27.93	14	Leader	0.66	10	0.74	15	Leader
Italy	ITA	34.24	27	Leader	0.64	11	0.94	5	Leader
Jamaica	JAM	56.72	59	Moderate	0.31	59	0.51	40	Follower
Japan	JPN	32.05	20	Leader	0.49	27	0.55	30	Follower
Jordan	JOR	57.85	60	Moderate	0.30	68	0.37	83	Moderate
Kazakhstan	KAZ	50.75	50	Moderate	0.33	50	0.49	42	Follower
Kenya	KEN	74.45	89	Moderate	0.25	93	0.35	97	Beginner
Kyrgyzstan	KGZ	68.89	80	Moderate	0.29	70	0.51	39	Follower

continued on next page
continued from previous page

		OWSQ Approach			BoD using OWSQ weights		BoD using flexible weights ≥ 4%		
Country	Code	Score	Rank Position	Category	Composite Indicator	Rank Position	Composite Indicator	Rank Position	Category
Latvia	LVA	40.32	32	Follower	0.40	36	0.46	50	Follower
Lebanon	LBN	61.51	68	Moderate	0.31	61	0.37	81	Moderate
Lesotho	LSO	84.29	102	Beginner	0.21	111	0.44	57	Follower
Lithuania	LTU	41.38	36	Follower	0.43	34	0.49	45	Follower
Luxembourg	LUX	33.37	24	Leader	0.52	25	0.72	17	Leader
Macedonia	MKD	50.31	48	Moderate	0.33	47	0.40	69	Moderate
Madagascar	MDG	86.36	106	Beginner	0.24	98	0.34	100	Beginner
Malawi	MWI	86.56	107	Beginner	0.23	104	0.34	101	Beginner
Malaysia	MYS	42.61	38	Follower	0.39	38	0.44	58	Follower
Mali	MLI	87.61	111	Beginner	0.20	116	0.32	113	Beginner
Malta	MLT	32.55	21	Leader	0.63	13	1.00	1	Leader
Mauritius	MUS	49.79	47	Moderate	0.32	53	0.38	78	Moderate
Mexico	MEX	49.47	46	Moderate	0.31	57	0.41	67	Moderate
Moldova	MDA	57.98	61	Moderate	0.30	69	0.39	74	Moderate
Mongolia	MNG	65.31	75	Moderate	0.26	86	0.48	46	Follower
Montenegro	MNE	45.79	43	Follower	0.34	46	0.40	70	Moderate
Morocco	MAR	63.41	71	Moderate	0.28	77	0.36	90	Beginner
Mozambique	MOZ	87.39	110	Beginner	0.22	107	0.38	80	Moderate
Namibia	NAM	73.51	87	Moderate	0.25	94	0.35	94	Beginner
Nepal	NPL	74.68	90	Moderate	0.26	84	0.36	89	Moderate
Netherlands	NLD	23.22	6	Leader	0.73	5	0.80	10	Leader
Nicaragua	NIC	73.10	85	Moderate	0.27	83	0.35	98	Beginner
Nigeria	NGA	86.10	105	Beginner	0.20	113	0.33	108	Beginner
Norway	NOR	19.54	3	Leader	0.73	6	0.84	7	Leader
Pakistan	PAK	79.06	97	Beginner	0.21	112	0.34	103	Beginner
Panama	PAN	51.06	51	Moderate	0.33	48	0.55	29	Leader
Paraguay	PRY	74.24	88	Moderate	0.26	90	0.31	117	Beginner
Peru	PER	56.08	57	Moderate	0.30	65	0.39	77	Moderate
Philippines	PHL	59.37	64	Moderate	0.31	60	0.51	35	Follower
Poland	POL	37.67	30	Follower	0.38	39	0.44	61	Moderate
Portugal	PRT	33.11	23	Leader	0.67	9	0.73	16	Leader
Romania	ROU	37.49	28	Follower	0.47	29	0.60	26	Leader
Russia	RUS	48.85	45	Moderate	0.32	51	0.46	52	Follower
Rwanda	RWA	75.84	94	Moderate	0.25	91	0.35	96	Beginner
Senegal	SEN	79.06	98	Beginner	0.25	95	0.38	79	Moderate
Serbia	SRB	40.42	33	Follower	0.43	32	0.51	37	Follower
Slovakia	SVK	33.82	26	Leader	0.54	24	0.77	11	Leader
Slovenia	SVN	29.20	17	Leader	0.61	17	0.69	21	Leader
South Africa	ZAF	65.34	76	Moderate	0.29	73	0.35	95	Beginner
South Korea	KOR	33.40	25	Leader	0.44	31	0.51	39	Follower
Spain	ESP	31.86	19	Leader	0.54	23	0.61	25	Leader
Sri Lanka	LKA	58.97	63	Moderate	0.31	56	0.53	32	Follower
Sweden	SWE	19.37	2	Leader	0.88	2	0.96	3	Leader
Switzerland	CHE	19.30	1	Leader	1.00	1	1.00	1	Leader
Tajikistan	TJK	76.93	95	Moderate	0.25	92	0.36	92	Beginner
Tanzania	TZA	82.79	101	Beginner	0.23	100	0.34	102	Beginner
Thailand	THA	43.68	39	Follower	0.34	43	0.44	62	Moderate
Tunisia	TUN	62.00	69	Moderate	0.30	64	0.37	82	Moderate
Turkey	TUR	52.88	53	Moderate	0.32	54	0.37	84	Moderate
Uganda	UGA	85.16	104	Beginner	0.23	99	0.34	104	Beginner
Ukraine	UKR	61.14	66	Moderate	0.29	71	0.40	72	Moderate
United Kingdom	GBR	25.78	10	Leader	0.64	12	0.67	22	Leader
United States	USA	28.22	15	Leader	0.58	20	0.76	12	Leader
Uruguay	URY	44.82	42	Follower	0.37	41	0.52	34	Follower
Venezuela	VEN	73.14	86	Moderate	0.28	78	0.40	73	Moderate
Vietnam	VNM	61.22	67	Moderate	0.28	74	0.36	86	Moderate
Yemen	YEM	93.90	118	Beginner	0.20	114	0.32	114	Beginner
Zambia	ZMB	82.20	100	Beginner	0.23	102	0.34	105	Beginner

Zimbabwe	ZWE	86.86	108	Beginner	0.22	106	0.33	107	Beginner

Appendix B - Composite indicators generated to analyse countries' performance within clusters

Table B.1 - Composite Indicator - Cluster of countries from Europe.

Country	Code	Composite Indicator	Rank Position
Albania	ALB	0.92	12
Austria	AUT	0.83	26
Belgium	BEL	0.88	17
Bosnia-Herzegovina	BIH	0.75	33
Bulgaria	BGR	0.85	24
Croatia	HRV	0.79	30
Czech Republic	CZE	0.86	21
Denmark	DNK	0.92	13
Estonia	EST	1.00	1
Finland	FIN	1.00	1
France	FRA	0.74	34
Germany	DEU	0.85	23
Greece	GRC	0.78	31
Hungary	HUN	0.86	19
Ireland	IRL	0.89	15
Italy	ITA	1.00	1
Latvia	LVA	0.72	36
Lithuania	LTU	0.73	35
Luxembourg	LUX	0.86	22
Macedonia	MKD	0.86	20
Malta	MLT	1.00	1
Moldova	MDA	0.82	27
Montenegro	MNE	0.81	29
Netherlands	NLD	0.90	14
Norway	NOR	0.98	8
Poland	POL	0.71	37
Portugal	PRT	0.93	11
Romania	ROU	0.98	9
Russia	RUS	1.00	1
Serbia	SRB	0.96	10
Slovakia	SVK	0.87	18
Slovenia	SVN	0.88	16
Spain	ESP	0.81	28
Sweden	SWE	0.99	7
Switzerland	CHE	1.00	1
Ukraine	UKR	0.75	32
United Kingdom	GBR	0.85	25

Table B.2 - Composite Indicator - Cluster of countries from Sub-Saharan Africa.

Country	Code	Composite Indicator	Rank Position
Angola	AGO	0.29	7
Benin	BEN	0.20	26
Botswana	BWA	0.48	3
Burkina Faso	BFA	0.27	15
Burundi	BDI	0.29	8
Cameroon	CMR	0.22	25
Cote d'Ivoire	CIV	0.25	20
Ethiopia	ETH	0.26	18
Ghana	GHA	0.28	13
Guinea	GIN	0.29	6
Kenya	KEN	0.27	16
Lesotho	LSO	0.29	5
Madagascar	MDG	0.25	21
Malawi	MWI	0.24	23
Mali	MLI	0.29	9
Mauritius	MUS	1.00	1
Mozambique	MOZ	0.25	19
Namibia	NAM	0.37	4
Nigeria	NGA	0.24	22
Rwanda	RWA	0.28	10
Senegal	SEN	0.28	11
South Africa	ZAF	1.00	1
Tanzania	TZA	0.28	12
Uganda	UGA	0.27	14
Zambia	ZMB	0.23	24
Zimbabwe	ZWE	0.26	17