



Combining reference point based composite indicators with data envelopment analysis: application to the assessment of universities

Samira El Gibari^{1,2} · Trinidad Gómez² · Francisco Ruiz²

Received: 27 April 2021 / Accepted: 2 June 2022 / Published online: 8 July 2022
© The Author(s) 2022

Abstract

In our knowledge society, where universities are key players, the assessment of higher education institutions should meet the new demands of the present complex environment. This calls for the use of techniques that are able to manage this complexity. In this paper, we propose a novel combination of methodologies, jointly using a multi-criteria reference point scheme and the data envelopment analysis (DEA) for the assessment of universities. This combination allows us to take into account all the aspects regarded as relevant to assess university performance, and use them as outputs in the efficiency analysis. Our findings highlight the convenience to assess the university performance by using both compensatory and non-compensatory schemes. This way, the information provided allows to detect the actions needed to improve the performances of the universities, rather than just giving an overall performance measure. Furthermore, combining the use of composite indicators with the DEA analysis provides a more complete picture of the institutions assessed, allowing universities to check their efficiency and to detect their weaknesses and strengths accordingly. The approach is illustrated using data of 47 Spanish public universities for the academic year, 2016–2017.

Keywords Multiple criteria analysis · Composite indicators · Compensation · Data envelopment analysis · Higher education system

Introduction

The evaluation of universities' performance has become indispensable, due to their vital role in the economic growth and social development of our society. Higher education institutions are key drivers of growth of performance, prosperity and competitiveness, at both the national and international levels and as a result, university rankings, an outcome of the

✉ Samira El Gibari
elgsamira@uma.es

¹ Programa de Doctorado en Economía y Empresa, Universidad de Málaga, Calle Ejido 6, 29071 Málaga, Spain

² Department of Applied Economics (Mathematics), University of Malaga, Calle Ejido 6, 29071 Málaga, Spain

competitiveness boosted by globalization, have put pressure on the universities to adapt to market forces without compromising quality (Heitor & Horta, 2013). However, most of the existing rankings operate under the concept of “the higher the value for all the indicators, the higher the performance achieved” (González-Garay et al., 2019). Hence, they assign certain performance measures to universities, regardless of whether they make an efficient use of their resources or not.

An increasing pressure on public finances is giving rise to the need to operate public universities with a higher degree of efficiency (Abbott & Doucouliagos, 2003). Since the pioneering works (Charnes et al., 1978, 1981), the comparative analysis of the efficiency in education has been brought into the focus of public and policy interest (De Witte & López-Torres, 2017). There is no doubt that this topic will continue to dominate headlines in the field of higher education, where a university is regarded as efficient if it produces the highest output levels by using the lowest input levels (Visbal-Cadavid et al., 2017). Specifically, following the economic recession in 2008, the general trend in reducing public funds for Spanish universities has led to define the allocation of resources in an efficient way as a priority for the Spanish public higher education sector (de la Torre et al., 2017a; Martínez-Campillo & Fernandez-Santos, 2020).

Data envelopment analysis (DEA) and its variations have been the most frequently used methodologies for measuring the efficiency in the context of higher education. Empirical research on the higher education efficiency using DEA models is becoming increasingly important at the international level (e.g. Abbott & Doucouliagos, 2003; Agasisti et al., 2019; Johnes, 2006). In Spain, according to Martínez-Campillo & Fernandez-Santos (2020), there has been limited research on the technical efficiency of the public universities (Bebegal-Mirabent et al., 2013; Bebegal-Mirabent, 2018; de la Torre et al., 2017a, 2017b; de Jorge Moreno et al., 2019; Martínez-Campillo & Fernandez-Santos, 2020; Salas-Velasco, 2020a, 2020b).

This approach is suitable in this context for several reasons: (a) it can handle multiple inputs and multiple outputs in a simple manner, (b) it does not assume any underlying functional relationships between inputs and outputs and (c) it determines the weights of inputs and outputs in an endogenous manner, looking for those that evaluate each unit in the best possible way, as compared to all the units (Abbott & Doucouliagos, 2003). However, the suitable number of inputs and outputs in relation to the number of units included in the analysis is still debatable (Akbari et al., 2020). Some authors claim that the number of units has to be at least twice the number of inputs and outputs to be used (Dyson et al., 2001), while others suggest that the number of units has to be three times higher. In fact, problems of discrimination between efficient and inefficient units often arise when there is a relatively large number of variables, as compared to the number of units (Charles et al., 2019) and thus, the greater the number of inputs and outputs to be used in relation to the number of units, the lower the discrimination in assessing relative efficiency (de Jorge Moreno et al., 2019). It is worth highlighting some existing approaches on variable (indicator) aggregation before DEA, such as principal component analysis (Adler & Yazhemsky, 2010), where the main applying components replace the original variables. It has also been suggested to set weights to the original variables, using information about prices (Nguyen & Zelenyuk, 2021), or value judgments via analytic hierarchical process (Meng et al., 2008), or applying a variant of DEA which offers the summarized variable from variable subsets defined by experts (Charles et al., 2019). In our case, the output variables in the DEA model will be composite indicators that summarize information on multiple aspects of the units' performance, in order to analyze whether better performance is jointed with an efficient use of available resources.

In the particular case of universities, most of the papers that use DEA combine labor and capital inputs to produce two main outputs: human capital (teaching) and scientific knowledge (research). Few studies, such as those by Berbegal-Mirabent et al. (2013) and de la Torre et al. (2017a) consider the technology transfer as a separated mission from the traditional ones of the university (teaching and research). We follow this focus in this paper. Interestingly, the number of outputs considered in the previously referenced papers ranges from two to four, being the number of graduate students and the number of publications the most frequently used ones. Nevertheless, the higher education system is quite diverse and multi-faceted, and therefore, it is difficult to condense the diversified activity going on within universities into few variables as outputs proxies (González-Garay et al., 2019). For this reason, it is necessary to introduce new metrics that allow us to combine a greater variety of aspects and avoid, at the same time, the use of an excessive number of variables, which could make the DEA model reduce its discriminatory power. Besides, in today's globalized and highly competitive environment, the higher education system should have a powerful tool to detect improvement areas, rather than just ranking the universities. Consequently, the evaluation of the university system requires the use of techniques that are able to manage all this complexity (Attardi et al., 2017).

In recent years, the measurement of such multidimensional frameworks has received special attention. Within this context, the use of composite indicators is highly suitable. They enable aggregation of many single indicators into one measure, making it possible to compare many units (Molinos-Senante, 2018). This way, they integrate large amounts of information in a clear and understandable way that is easy for decision makers to interpret (Szuwarzynski, 2018). The use of composite indicators greatly simplifies the comparison among higher education institutions, helping them to benchmark their performance against each other in a straightforward manner (González-Garay et al., 2019).

Few initiatives have tackled the evaluation of university performance without using composite indicators, such as the *U-Multirank*, which considers the scores of universities on individual indicators and place these in five performance groups (“very good” through to “weak”). Oppositely, a general widespread trend towards the construction of composite indicators has been established in the field. Most of the existing rankings have developed composite indicators to provide rankings of higher education institutions or countries (the *ARWU*; the Times Higher Education World University Ranking (*THE*); the QS World University Ranking (*QS*); the *U-Ranking* project of the BBVA Foundation and the *IVIE*¹ (Aguillo et al., 2008; Alasehir et al., 2014; Giannoulis & Ishizaka, 2010; Torres-Salinas et al., 2011).

The construction of composite indicators has been tackled from different angles. One of the most extended approaches consists of using multiple criteria decision making (MCDM) methods (see, e.g., El Gibari et al., 2019). Murias et al. (2008) present a composite indicator for quality assessment in the Spanish public university system using DEA, while Giannoulis and Ishizaka (2010) adopt ELECTRE III to evaluate the performance of British universities, reflecting personal preferences. El Gibari et al. (2018) develop composite indicators for different compensation degrees for each of the three main missions of universities (research, teaching and technology transfer). To this end, they evaluate the performance of public universities using a multi-criteria analysis technique, based on the double reference point method (Ruiz et al., 2011; Wierzbicki, 1980). The use of this technique in the evaluation of university performance has two main advantages. First, the use of reference levels for each indicator makes

¹ The Valencian Institute of Economic Research.

the results obtained easily interpretable in terms of the performance of the university with respect to these levels. Second, aggregations for different compensation degrees are provided, and thus, apart from giving an overall performance measure of the universities, they also provide warning signals that assists the user in strategic decision making for policy purposes. This method was generalized to the multiple reference point weak-strong composite indicator (*MRP-WSCI*) approach, where any number of reference levels and any scale can be used (Ruiz et al., 2020).

According to Nardo et al. (2005), weighting is a key step in constructing composite indicators. Weights can have a significant effect on the overall composite indicator and the results obtained. The literature provides a wide range of methodological approaches. Generally, weights are usually adopted on the basis of expert's opinions and decisions makers, but the equal weighting approach is also widely used (Antanasijevic et al., 2017; Blancas et al., 2010; Langhans et al., 2014; Zhou et al., 2010). Anyway, weighting implies introducing controversial and subjective elements on which there is no general consensus (Huang, 2011; Marginson & Van der Wende, 2007).

In this paper, we discuss how the combination of the *MRP-WSCI* approach with the data envelopment analysis (DEA) can add value to traditional composite indicators, not only in terms of securing high performance, but also in how to make an appropriate use of resources. Thus, the contribution of this paper to the current literature is twofold. First, we analyze the added information provided by the joint consideration of the *MRP-WSCI* composite indicators and the DEA analysis. Unlike the approach followed in previous studies, we use different composite indicators for each mission (teaching, research and technology transfer), which allow us to take into account a greater variety of aspects regarding the activities of universities. Besides, we consider two scenarios, depending on whether the compensability among certain indicators is or not possible, in order to test its influence on the evaluation of the efficiency of universities. Second, the DEA efficiency scores for the Spanish public universities is carried out in a robust environment, with respect to changes of the weights when constructing the *MRP-WSCI* composite indicators. The suggested methodology is applied to the case of the Spanish public universities, but it can be used as a supporting tool in other university systems, allowing decision makers to detect and improve important factors of the educational efficiency and effectiveness of universities. In fact, the ultimate goal of the research proposed in this article is to provide the decision makers of the higher education system with a tool that allows them to evaluate the performance and efficiency of the universities and make decisions accordingly, taking into account their preferences.

Following this introduction, “[Available system of indicators](#)” gives a short overview of the data used for the evaluation of the Spanish public universities. “[Research methodologies](#)” describes the *MRP-WSCI* procedure and its combination with the DEA analysis. The empirical results for the compensatory and non-compensatory scenarios are presented in “[Compensatory scenario](#)” and “[Non-compensatory scenario](#)”, while “[Discussion](#)” discusses some remarkable aspects related with the results of this paper, and finally, “[Conclusions](#)” presents some conclusions.

Available system of indicators

The first step of our study consists of analyzing the performance of the Spanish public universities. To this end, we have used an existing set of indicators that is frequently used by several Spanish institutions. More precisely, a set of 23 performance indicators is selected,

taking into account both their relevance to measure the Spanish public universities performance and their availability (for further details about the performance indicators considered in each mission, see Table 1). All of them are outcome indicators, and some have been relativized in order to avoid biases due to the size of the university (Alasehir et al., 2014). We would like to point out that the aim of this research is showing the validity of the methodology proposed in this field. We do not discuss the suitability of the particular system of indicators used, which is the system used by Spanish higher education managers. Our methodology can be applied to any other system of available indicators, should it be regarded as more suitable by the decision makers.

The research and teaching missions are further broken down into three and two sub-blocks, respectively, while the technology transfer missions has no sub-blocks. We have considered that indicators within a sub-block are substitutable, that is, poor performance in some ones can be offset by good behaviors in other ones. However, the different sub-blocks belonging to the same mission may be substitutable or not. This issue will lead to two scenarios depending on whether the compensability is assumed or not.

Second, we measure the efficiency of the Spanish public universities, where the outputs considered are the composite indicators provided by the *MRP-WSCI* approach. According to (Salas-Velasco, 2020b), the “perfect set of inputs” to evaluate university efficiency does not exist. In fact, the literature of higher education efficiency provides a wide range of inputs. According to Berbegal-Mirabent et al. (2013) and Anderson et al. (2007), in the higher education field, the labour force and capital inputs are vital. In particular, it is important to consider faculty members, as they transmit knowledge to students and carry out most of the research, while the R & D income provides financial resources necessary to develop new research and technology transfer outputs. Moreover, certain financial expenditures are key inputs to the development of the activities of each university (Zhong et al., 2011). In this paper, four inputs are considered: *Ratio of faculty members with PhD*, *R & D funds per faculty member*, *Staff expenses per students* and *Expenses on good and services per student* (similar to, Agasisti & Pérez-Esparrells, 2010; Expósito-García & Velasco-Morente, 2018; Yang et al., 2018, these expenses refer mainly to expenditure on academic staff, expenditure on non-academic staff and running expenses in relation to goods and services). The input and output specifications of each DEA analysis performed are displayed in Table 2.

In summary, the performance indicators and the inputs considered in this paper have been used by the Spanish Education authorities to evaluate the Spanish universities in different areas, such as access to funding, quality and excellence, internationalization and results. Our database contains data on research, teaching and technology transfer for 47 Spanish public universities for the academic year 2016–2017 (Table 5 in the Appendix records the names of such universities and their abbreviations). The data were collected from two sources: the Spanish University Rectors’ Conference² (*CRUE*) and the *IUNE* Observatory³.

² <http://www.crue.org/Publicaciones/Paginas/UEC.aspx?Mobile=0>.

³ <http://www.iune.es>

Table 1 System of performance indicators in each mission

Mission	Sub-block	Individual indicator	Description	Weight
The MRP- WSCI approach	Research	Publications per doctor	Universities' yearly scientific output/number of full-time equivalent (FTE) faculty members holding a PhD	0.31
		Citations per doctor	Total citations received/number of FTE faculty members holding a PhD	0.31
	Projects (0.35)	Q1 publications per doctor	Total papers in first quartile/number of FTE faculty members holding a PhD	0.38
		Training grants and contracts per doctor	Number of training grants and contracts (including all national plans)/number of FTE faculty members holding a PhD	0.23
		National projects assigned per doctor	Number of national projects awarded/number of FTE faculty members holding a PhD	0.38
	Other research activities (0.23)	European projects assigned per doctor	Number of European projects awarded/number of FTE faculty members holding a PhD	0.38
		International collaboration	Papers involving international collaboration/Universities' yearly scientific output	0.24
		Theses defended per doctor	Total dissertations defended/number of FTE faculty members holding a PhD	0.16
	Teaching	Participation in projects	(Number of faculty members who participate in European and national competitive scientific projects/Total number of faculty members)*100	0.36
		Official recognition research	(Number of six-year merits awarded to the university/Number of possible sixyear merits)*100	0.24
Ratio of postgraduated student		Number of postgraduated students/total number of students	0.18	
Success rate		Number of course credits passed per student/number of course credits assessed per student	0.23	
		Assessment rate	Number of course credits assessed per student/number of course credits enrolled per student	0.18

Table 1 (continued)

Mission	Sub-block	Individual indicator	Description	Weight
		Graduation rate	(Number of students, among those newly enrolled in a cohort, who pass, in the expected time or in one more year, the credits needed to obtain the degree/Total number of new students of the same cohort of the corresponding university degree)*100	0.23
		Drop-out rate	(Number of students belonging to the entry cohort that first enrolled, in the first course, in the academic year 2015–2016 and who left without academic accreditation, not enrolling in the courses 2015–2016 and 2016–2017/Number of students in the cohort who first enrolled in the academic year 2015–2016)*100	0.18
	External projection (0.45)	International attractiveness	Number of undergraduate students coming from other countries/total number of students	0.425
		National students in an Erasmus program	Number of Spanish undergraduate students who participate in Erasmus/total number of students	0.189
		Foreign students in an Erasmus program	Number of students coming from other countries participating in Erasmus/total number of students	0.232
		Master degrees in a foreign language (or bilingual)	Number of bilingual official master's degrees/total number of master's degrees	0.154
	Technology transfer	National patents per doctor	Number of national patents registered/number of FTE faculty members holding a PhD	0.29
		R & D contracts and services rendered per doctor	Number of R & D and consultancy agreements/number of FTE faculty members holding a PhD	0.29
		PCT extensions per doctor	Number of international patents (PCT) extensions/number of FTE faculty members holding a PhD	0.23
		Spin-offs per doctor	Number of spin-offs/number of FTE faculty members holding a PhD	0.19

Table 2 Input and output specifications of the DEA analysis performed

DEA analysis performed		Input and output variables	Description
The DEA analysis	Compensatory scenario	Inputs	Staff expenses per student
			Expenses on good and services per student
			R & D Funds per faculty member
			Ratio of faculty members with PhD
	Noncompensatory scenario	Outputs	MRP-WSCI composite indicators
		Inputs	Staff expenses per student
			Expenses on good and services per student
			R & D funds per faculty member
			Ratio of faculty members with PhD
		Outputs	MRP-WSCI composite indicators
		Total staff expenditure (chapter 1 of the liquidated budgets)/total number of students	
		Total expenses on good and services (chapter 2 of the liquidated budgets)/total number of students	
		Budgety received from R &D/number of FTE faculty members (€ per faculty members)	
		Number of FTE faculty members holding a PhD/number of FTE faculty members	
		3 outputs (WSCI of each mission)	
		Total staff expenditure (chapter 1 of the liquidated budgets) / Total number of students	
		Total expenses on good and services (chapter 2 of the liquidated budgets)/total number of students	
		Budgety received from R &D/number of FTE faculty members (€ per faculty members)	
		Number of FTE faculty members holding a PhD/number of FTE faculty members	
		3 outputs (SWCI of each mission)	

Research methodologies

In this paper, we propose a combination of the *MRP-WSCI* approach (Ruiz et al., 2020) with the DEA analysis (Charnes et al., 1978) for the evaluation of the Spanish public universities. This section describes both methodologies. Then, we describe the stages for combining them.

MRP-WSCI approach

The *MRP-WSCI* approach, defined in Ruiz et al. (2020), is an adaptation of the original reference point method proposed by Wierzbicki (1980) to the construction of composite indicators, which generalizes the double reference point method (Ruiz et al., 2011; Wierzbicki et al., 2000). In order to construct the *MRP-WSCI* composite indicators, different aspects are taken into account:

- Reference levels. The performance of the units (J), in our case, 47 Spanish public universities, is assessed in terms of how far is each indicator (i) from given reference levels. The decision-maker can give any number of reference levels (q_i^t) for each indicator. These reference levels can be “absolute” or “relative”. In the first case, the reference levels are given by one or a group of experts, if they have enough knowledge about the problem and they wish to do so. In this case, the final composite indicator gives us an absolute measure of performance, with respect to these values. Alternatively, they can be set statistically, given a data set. In this case, the composite indicator measures the relative position of the units with respect to those belonging to the data set.
- Achievement functions. Once the reference levels are established, a so-called achievement function ($s_i(x_{ij}, q_i)$) measures the deviation between the values of the objective functions and the reference levels, and at the same time, brings all the indicators down to a common scale. The achievement function for the “the more, the better” type of indicators is calculated as follows:

$$s_{ij} = s_i(x_{ij}, q_i) = \alpha^{t-1} + \frac{\alpha^t - \alpha^{t-1}}{q_i^t - q_i^{t-1}}(x_{ij} - q_i^{t-1}) \quad \text{if } x_{ij} \in [q_i^{t-1}, q_i^t], \quad (t = 1, \dots, n + 1), \tag{1}$$

where x_{ij} is the value of indicator i ($i = 1, \dots, I$) for unit j ($j = 1, \dots, J$), q_i^t ($t = 0, \dots, n + 1$) are the reference levels of indicator i and each α^t ($t = 0, \dots, n + 1$) is the value in the common scale.

- Weights. Unavoidably, the different indicators used to measure university performance do have different relative importance, and this has to be taken into account when building an overall measure. The methodology must be designed in order to clearly identify these subjective elements, and to allow potential users to include their own weights (μ_i) in the process.
- Compensation. The *MRP-WSCI* method allows to construct two different composite indicators, depending on the compensation degree among the indicators.

First, the weak indicator (*WCI*) allows for full compensation among the individual indicators.

$$WCI_j = \sum_{i=1}^I \mu_i^w s_i(x_{ij}, q_i). \quad (2)$$

where μ_i^w are the normalized weights, which add up to 1.

Second, the strong indicator (*SCI*) does not allow for any compensation. Therefore, it reflects the worst values achieved by an unit, relativized by the weight of the indicator. To this end, the original achievement functions have to be modified as follows:

$$\bar{s}_i(x_{ij}, q_i) = \begin{cases} \alpha^t + (s_i(x_{ij}, q_i) - \alpha^t)\mu_i^s, & \text{if } s_i(x_{ij}, q_i) \in (\alpha^{t-1}, \alpha^t], \\ & (t = 1, \dots, n+1), \\ \alpha^1 + (\alpha^0 - \alpha^1)\mu_i^s, & \text{if } s_i(x_{ij}, q_i) = \alpha^0. \end{cases} \quad (3)$$

where μ_i^s are the corrected normalized weights. When constructing the *SCI*, a different normalization of the weights is carried out, where the greatest weight takes value 1. This way, *SCI* takes a worse value if the corresponding indicator has a higher weight, and in particular, it takes the worst possible value if unit j has the worst possible value in the highest weighted indicator.

Therefore, the *SCI* takes the form:

$$SCI_j = \min_{i=1, \dots, I} \{\bar{s}_i(x_{ij}, q_i)\}. \quad (4)$$

Summing up, the *WCI* gives an overall performance measure of the units (universities), while the *SCI* provides warning signals that let the user detect improvement areas.

Data envelopment analysis

Data envelopment analysis (DEA) is a non-parametric technique, originally introduced by Farrell (1957) and extended by Charnes et al. (1978, 1981). It was proposed to analyse the efficiency problem of complex organizations that produce many outputs ($o = 1, \dots, O$) using several inputs ($m = 1, \dots, M$) at a time.

In a DEA model, technical efficiency is defined as the relative ability of each decision making unit (DMU, $j = 1, \dots, J$) in producing outputs given a certain set of inputs. An important aspect to emphasize in the DEA model is the choice of a set of weights which combines several outputs and several inputs. This issue is left to DEA through a linear programming technique which chooses the best set of weights for each DMU to maximize the efficiency ratio (outputs/inputs).

In this paper, we employ an output-oriented model, because our aim with the DEA model is to analyze whether the performance of a university, measured by MRP-WSCI indicators (outputs), is the best possible one, given its available resources (inputs). Besides, we use the BCC model (Banker et al., 1984), which assumes variable returns to scale (VRS), so that each university is compared to others of the same “relative” size. Both hypotheses are usual in this context (Agasisti & Pérez-Esparrells, 2010; Visbal-Cadavid et al., 2017). Furthermore, in our case, the BCC model allows us to guarantee that the projected output is within the common scale used for the composite indicators obtained in the first step.

Given these considerations, the linear program used to obtain the efficiency level of each DMU (university j) is:

$$(P_j) \left\{ \begin{array}{l} \max_{\phi_j, \lambda_k, s_o^+, s_m^-} \quad \phi_j + \varepsilon \left(\sum_{o=1}^O s_o^+ + \sum_{m=1}^M s_m^- \right) \\ \text{subject to} \quad \sum_{k=1}^J \lambda_k x_{mk} + s_m^- = x_{mj}, \quad (m = 1, \dots, M) \\ \sum_{k=1}^J \lambda_k y_{ok} - s_o^+ = \phi_j y_{oj}, \quad (o = 1, \dots, O) \\ \sum_{k=1}^J \lambda_k = 1 \\ \lambda_k, s_o^+, s_m^- \geq 0, \end{array} \right. \tag{5}$$

where $1 \leq \phi_j < +\infty$ indicates the score efficiency for university j . When $\phi_j = 1$, university j attains an efficient (best-practice) performance, in the sense that each of its outputs could not be improved, remaining the inputs at the same level. Contrariwise, when $\phi_j > 1$, university j is failing to attain best-practice performance. λ_k is the intensity vector and ε is the non-Archimedean infinitesimal constant. s_o^+ is the output slack vector, while s_m^- is the input slack vector. y_{oj} is the amount of output o from university j and x_{mj} is the amount of input m from university j .

Combining the *MRP-WSCI* approach and the DEA analysis

In order to evaluate the Spanish public universities by combining the *MRP-WSCI* approach and the DEA analysis, as can be seen in Fig. 1, the empirical analysis carried out in this paper follows these steps:

1. Construction of the *MRP-WSCI* composite indicators for each mission, in a robust environment with respect to the indicator weights.
 - Normalization. In this paper, in order to compare the Spanish public universities among themselves, the statistical option for establishing the reference levels has been chosen. This way, we will obtain a measure of the relative performance of each university with respect to all the Spanish public universities. Namely, for each indicator i , apart from the minimum (Min_i) and maximum (Max_i) values, we have used two intermediate reference levels: the reservation level (r_i , which corresponds to the average value between the mean and the Min_i value) and the aspiration level (a_i , which corresponds to the average value between the mean and the Max_i value). Once the reference levels are given, all the performance indicators are translated to a common default scale from 0 to 3, that is, we set $\alpha_0 = 0$, $\alpha_1 = 1$, $\alpha_2 = 2$ and $\alpha_3 = 3$, which define the performance levels of each indicator i , e.g. poor (0 to 1, the university performs worse than the corresponding r_i), fair (1 to 2, the university performs better than the r_i level, but worse than the a_i level), and good (2 to 3, the university performs better than the corresponding a_i). It should be mentioned that an outlier detection, using the interquartile range ($IQR = Q_3 - Q_1$) method was carried out (Tukey, 1977). This test allows to distinguish among mild outliers ($Q_1 - 1.5 \times IQR$, $Q_3 + 1.5 \times IQR$) or extreme outliers ($Q_1 - 3 \times IQR$, $Q_3 + 3 \times IQR$). In our case study, no extreme outliers were detected, while the mild outliers detected were assigned, on the corresponding

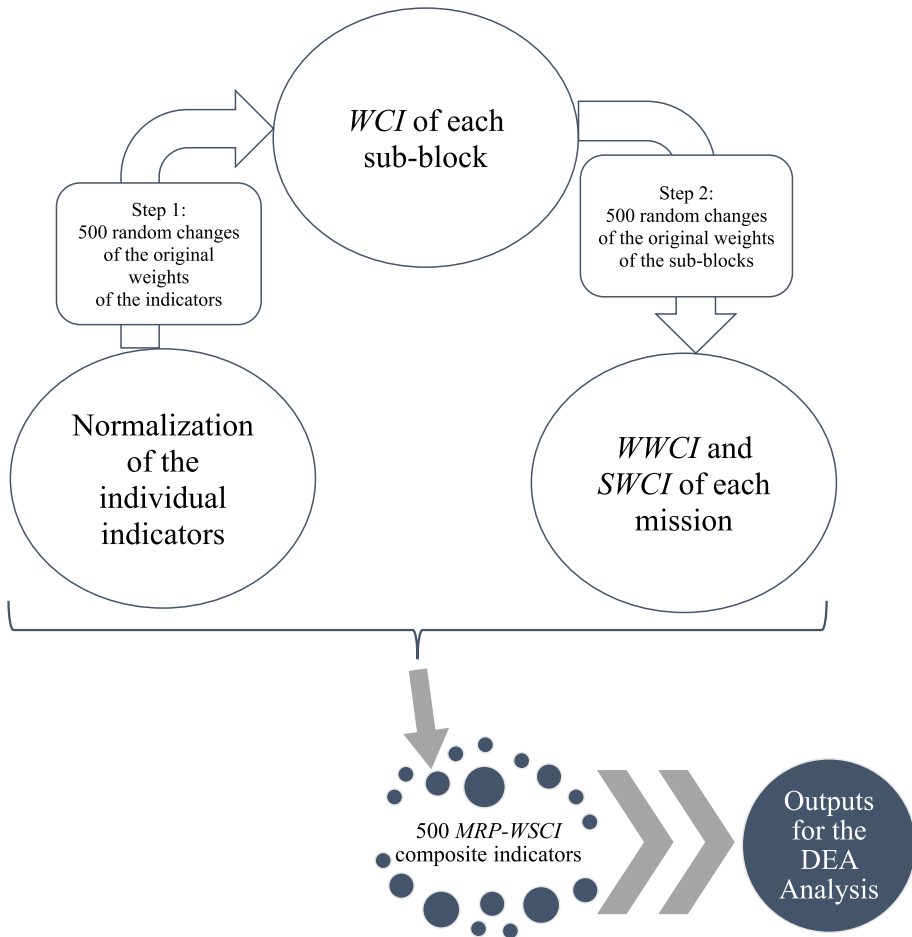


Fig. 1 Stages for combining the *MRP-WSCI* approach and the DEA analysis

achievement function, the maximum or minimum value (depending on its relative position) of the scale.

- **Weighting.** As previously mentioned, the developed methodology aims to provide decision makers with a useful tool for decision making. Therefore, the evaluation of the performance of the universities has to take into account their preferences, that is, the relative importance they give to the different elements of the system of indicators used. For this reason, in this study, the weights of the performance individual indicators and sub-blocks were assessed by a group of experts in the field of the Spanish universities, integrated by researchers and professionals from different fields of research and different Spanish public universities (the weights assigned to each indicator and sub-block are displayed in Table 1 in “[Available system of indicators](#)”). These experts have been chosen according to the criteria of people with great decision-making capacity in the field of the Spanish public university system. In any case, given the impact of these weights on the

final results obtained, in “[Discussion](#)” a comparison is made with the results that would be obtained if equal weights were used.

- Robustness analysis. In order to construct the *MRP-WSCI* composite indicators in a robust environment with respect to the indicator weights, we have implemented 500 computation rounds, changing randomly $\pm 10\%$ the weights of all the indicators (and corresponding sub-blocks) considered for the construction of the *MRP-WSCI* approach for each mission. The values of the weights are sampled from a uniform distribution centered in its actual value, and censored at 0 and 1.
- Aggregation. Finally, the *MRP-WSCI* composite indicators are calculated in a robust environment with respect to the indicator weights. In our case, as mentioned in “[Available system of indicators](#)”, the research and teaching missions are further broken down into three and two sub-blocks, respectively, while the technology transfer mission has no sub-blocks (for further details, see [Table 1](#) in “[Available system of indicators](#)”). This way, the aggregation is carried out in two steps:
 - In the first step, since the information provided by the individual indicators belonging to each sub-block is homogeneous, a full compensation among them is allowed. That is, bad performances in certain single indicators can be offset by good performances in other. This way, we get the *WCI* of each sub-block of the research and the teaching missions.
 - In the second aggregation step, different compensation degrees among the sub-blocks are provided. First, we construct the *WWCI*, which allows for a full compensation among the different sub-blocks. This way, the *WWCI* provides a measure of overall performance of each Spanish public university in each mission. Second, the *SWCI* does not allow for any compensation among the sub-blocks. Therefore, the use of *SWCI* is useful to point out the worst sub-block of each mission, and, to detect possible improvement areas of the Spanish public universities. In the case of the technology transfer mission, which has no sub-blocks, the *WCI* and *SCI* are obtained directly in the first step (by aggregating the individual indicators).

As mentioned before, the construction of the *MRP-WSCI* composite indicators is made in a robust environment with respect to the indicator weights. This way, we obtain 500 *MRP-WSCI* composite indicators for each compensatory scenario (and for each mission), that is, we construct 500 *WWCI* and 500 *SWCI* for each mission, changing randomly $\pm 10\%$ the corresponding weights considered.

2. The *MRP-WSCI* composite indicators (used later on as outputs in the DEA analysis) are represented through box-plots for the 500 instances for each Spanish public university (and for each mission). This way, we can graphically see the variability of the corresponding weak and strong composite indicators across the 500 instances, and analyze their reliability.
3. Once the performance of the Spanish public universities is analyzed using the *MRP-WSCI* approach, the next step consists of calculating the efficiency scores of the Spanish public universities, considering the composite indicators obtained as outputs. To this end, we perform two DEA analyses with different outputs. First, the *WWCI* of each mission are considered as outputs. Similarly, for each mission, we consider as outputs the *SWCI*. As mentioned in “[Available system of indicators](#)”, four different inputs are considered (*Ratio of faculty members with PhD*, *R & D funds per faculty member*, *Staff expenses per students* and *Expenses on good and services per student*). As previously

mentioned, all the composite indicators obtained are on a scale from 0 to 3, and since we use the BCC model (5), we have

$$\phi_j y_{oj} \leq \sum_{k=1}^J \lambda_k y_{ok} \leq 3,$$

that is, the projected output is in the same scale as the composite indicators. It must be noticed that a DEA analysis is carried out for each one of the 500 instances, and the robustness of the results is also analyzed.

4. Finally, the input and output levels and the efficiency scores are represented through bar charts for each Spanish public university.

Application to the Spanish public higher education system

In this section, we show and analyze the results obtained. To this end, the combination of the *MRP-WSCI* approach and DEA analysis developed in “[Combining the MRP-WSCI approach and the DEA analysis](#)” is applied to the data of the 47 Spanish public universities. As previously mentioned, different DEA analyses with different outputs are performed (for further details about the input and output specifications, see Table 2 in “[Available system of indicators](#)”).

First, we analyze the performance of the Spanish public universities in the compensatory scenario (*WWCI* of each mission), and we measure the efficiency of the universities in this scenario. Similarly, we analyze the performance and efficiency of the Spanish public universities in the non-compensatory scenario (*SWCI* of each mission). As mentioned in “[Available system of indicators](#)”, the inputs are the same in both analyses. Note that, as mentioned before, our analysis is made in a robust environment with respect to the performance indicator (and sub-blocks) weights.

Compensatory scenario

The results of the 500 instances for the compensatory scenario are shown in Fig. 2, where each sub-figure presents the box-plot for the 500 values of the *WWCI* for each Spanish public university, by changing randomly $\pm 10\%$ the weights of all the individual indicators (and corresponding sub-blocks) considered in each mission. We can observe that the range of variation in all missions is very small. In general, we can assume that the *WWCI* of the Spanish public universities are fairly robust, since the solution proved to be stable for all missions.

Let us analyze the overall performance of the Spanish public university system in each mission. As seen in Fig. 2, if we consider the median value of the *WWCI* for the 500 instances, most universities are in the “fair” performance level for the three missions ($1 < \text{value of the } WWCI < 2$). That is, most of the Spanish public universities perform better than the corresponding reservation level, but worse than the aspiration level for all the university missions analyzed. Note that the Spanish public universities are compared among themselves.

Specifically, in the research mission, among the 47 Spanish public universities in the data set, there is only one university (*ULPGC*) in the “poor” performance level (median value of the *WWCI* = 0.88 < 1), while four universities (*UPF*, *UAB*, *UAM* and *UB*) are

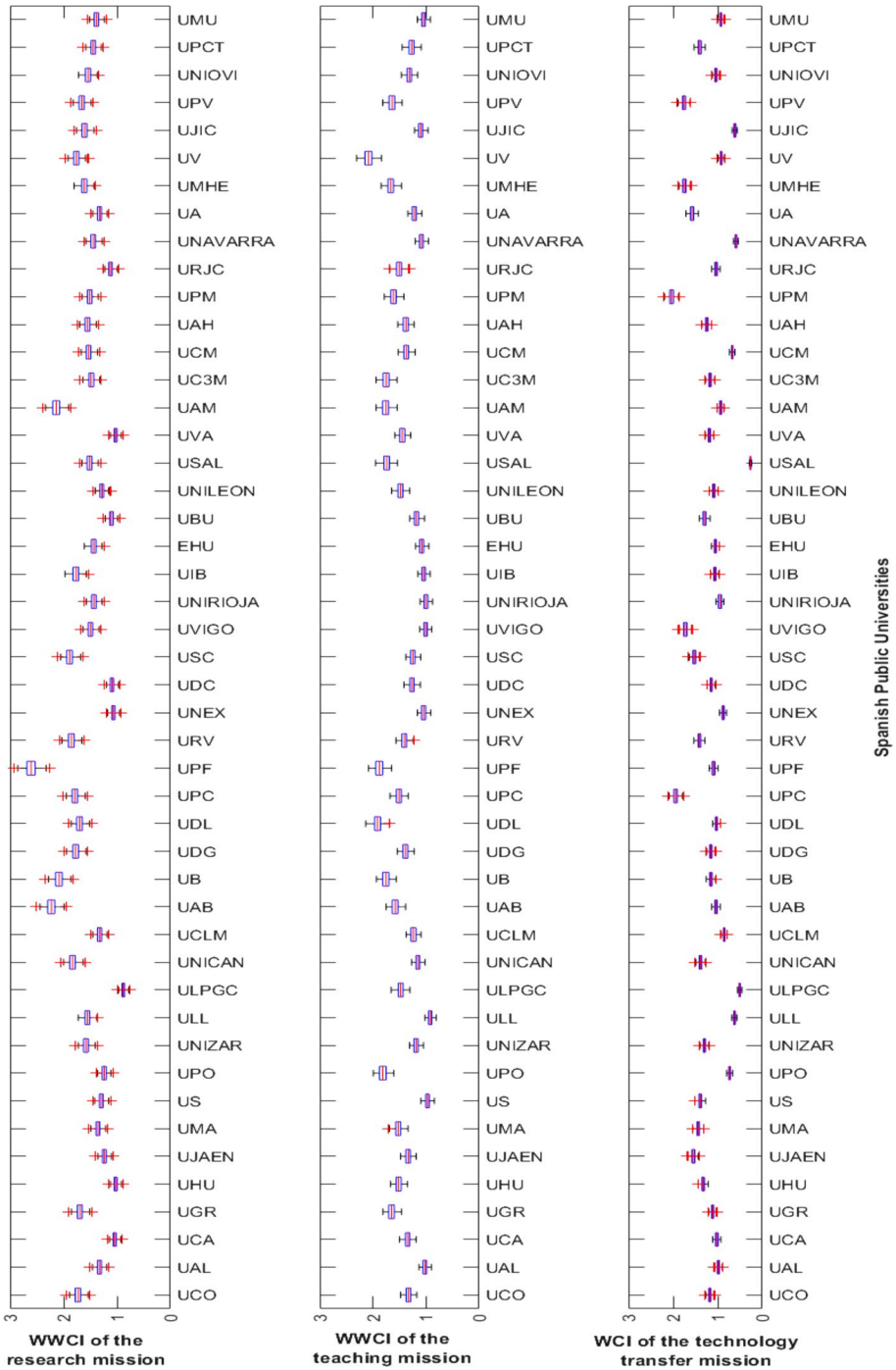


Fig. 2 Robustness analysis on weights: effect on WWCi of each mission

in the “good” performance level (median value of the *WWCI* > 2). The case of *UPF* is noteworthy, getting the best *WWCI* value for the 500 instances (median value of the *WWCI* = 2.60). This is because the “intermediate” performance in two individual indicators (*participation in projects* and *Official recognition research*) is compensated by good results in the rest of the research performance indicators. In fact, *UPF* has the best possible value (3), that is, the best value among all the Spanish universities, in four research indicators, while the rest of indicators are over the aspiration level. Regarding the teaching mission, in general, the overall performance of all the Spanish public universities is quite similar, and only one university (*UV*) achieves a median value of the *WWCI* better than the aspiration level (median value of the *WWCI* = 2.08 > 2), while two universities (*ULL* and *US*) have the median value of the *WWCI* worse than the reservation level (median value of the *WWCI* < 1). The rest of Spanish public universities are in the “fair” performance level (1 < value of the *WWCI* < 2). With respect to the technology transfer mission, a higher number of universities perform poorly. Specifically, the median values of the *WWCI* for fourteen Spanish public universities are worse than the reservation level, while only *UPM* performs better than the aspiration level in the technology transfer mission (median value of the *WWCI* = 2.04 > 2).

Next, we apply DEA to each of the 500 instances. First, it should be pointed out that the efficiency scores for all the Spanish public universities are also very robust, since the solution proved to be stable for the 500 instances, specially for the efficient ones ($\phi_j = 1$ for the 500 instances in all of them). Second, in order to analyze the DEA results in a more intuitive way, for each Spanish public university, Fig. 3 shows the input levels (the first four sub-figures), while the output levels (the fifth sub-figure) and the efficiency scores (the sixth sub-figure) are represented taking into account the median value of the 500 instances.

As can be seen in Fig. 3, on the one hand, 18 Spanish public universities are efficient in the compensatory scenario (colored in red in the bottom part of Fig. 3). That is, when a total compensation is allowed among the individual indicators (and the corresponding sub-blocks), the DEA analysis point towards the presence of a group of 18 Spanish universities that make an efficient use of their resources. The rest of the universities are not efficient (colored in blue). It must be taken into account that the level of inefficiency increases as the DEA inefficiency score grows. For example, a DEA score of 1.42 means that this university would have to significantly increase (42%) its level of outputs (*WWCI* of each mission), given its available inputs. For some of the most inefficient universities, it can be observed in Fig. 3 that, despite having high values of the inputs considered (topmost part of the figure), they do not manage to perform well in any of the three missions analyzed (middle part of the figure, research in blue, teaching in red and transfer in yellow).

Furthermore, our analysis shows that, in addition to the Spanish public universities that are always present in the most prestigious university rankings (e.g. *UPF*, *UAB*, *UAM* and *UB*), other Spanish public universities, occupying the lowest positions in some missions in the previous analysis, are among the efficient ones. Let us analyse, for example the case of *ULPGC*, which in Fig. 2, has the worst median value of the *WWCI* for the research mission (0.88) and the second worst median value of the *WWCI* for the technology transfer mission (0.51), while in the teaching mission, it performs better (1.48). However, the result of the DEA analysis (Fig. 3) points to an efficient use of its resources. It should be noted that *ULPGC* has the lowest value of *R & D funds per faculty member*. In other words, taking into account that the inputs available are quite modest, the performances are, comparatively, good enough. On the other hand, it is noteworthy the case of *UV*, which is among the efficient Spanish public universities.

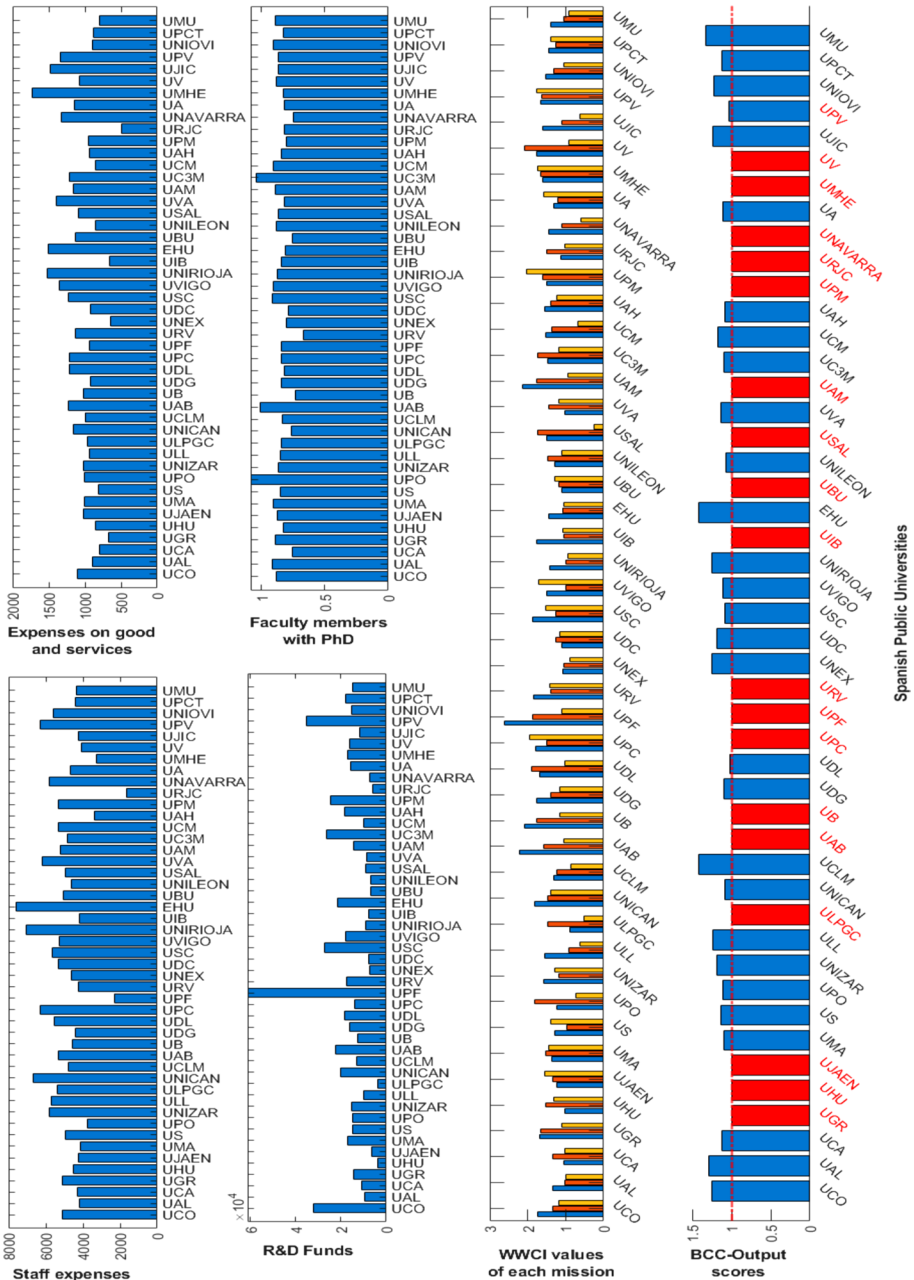


Fig. 3 Inputs, outputs and BCC-output scores of the and WWCI of the three missions

This university, despite having one of the lowest *Staff expenses per students*, was the only university with a median value of the WWCI for the teaching mission better than the aspiration level, and it has one of the best performances in the “fair” level for the

research mission (median value of the $WWCI = 1.76$), while it performs poorly in the technology transfer mission (median value of the $WWCI = 0.92$).

Therefore, these findings highlight the need to complement the use of composite indicators and university rankings with other analytic tools, such as the DEA analysis. Overall, according to González-Garay et al. (2019), DEA provides further insight into the assessment of universities, allowing institutions to better understand their weaknesses and strengths.

Non-compensatory scenario

Next, we will analyze the performance and efficiency of the Spanish public university system in the non-compensatory scenario ($SWCI$ of each mission).

Similarly to the previous analysis, the results of the 500 instances for the non-compensatory scenario are shown in Fig. 4, where each sub-figure presents the box-plot for the 500 values of the $SWCI$ for each Spanish public university, by changing randomly $\pm 10\%$ the weights of all the individual indicators (and corresponding sub-blocks) considered in each mission. First, comparing Figs. 2 and 4, we can notice that, to some extent, due to its compensatory nature, results for $WWCI$ are more robust than results for $SWCI$. Anyway, the results are still robust enough to be analyzed and used in the DEA analysis.

In this case, most of the Spanish public universities are in the “fair” performance level for the research and teaching missions ($1 < \text{values of the } SWCI < 2$), meaning that all their sub-blocks perform better than the reservation level. However, in the technology transfer mission, most universities are in the “poor” performance level (values of the $SCI < 1$). That is, most Spanish public universities perform worse than the corresponding reservation level for at least one indicator (given that there are no sub-blocks in this mission).

Let us analyze the results of the $SWCI$ for the research mission (Fig. 4, first sub-figure). Let us point out the case of *UPF*, which is the only Spanish public university in the “good” performance level (median value of the $SWCI = 2.43 > 2$, that is, with all the sub-blocks performing better than the aspiration level), and the one with a widest variability range compared to the rest of universities. First, *UPF* has the best possible value (3) in all the individual indicators for the *publications* sub-block, which has the highest weight in the research mission (0.425). This way, the median value of the WCI of the *publications* sub-block is 3. On the other hand, the median values of the WCI of the *projects* sub-block and the *other research activities* sub-block are 2.53 and 2.01, respectively (the sub-block weights are 0.345 and 0.23, respectively). The median value of the $SWCI$ of the research mission is 2.43. However, the $SWCI$ value for many experiments is worse than this median value (wider variability), which is mainly due to its worse performance in the *other research activities* sub-block (in the previous analysis, we have mentioned that this university has two individual indicators in the “intermediate” performance level, *participation in projects* and *Official recognition research*, which both correspond to the *other research activities* sub-block). As a result, when the weight of this sub-block is increased, the $SWCI$ gets a worse value. This allows decision makers at *UPF* learn that a special attention should be paid to these indicators.

With respect to the technology transfer mission, which has no sub-blocks, the SCI is obtained directly by aggregating the individual indicators. Only 8 Spanish public universities are in the “fair” performance level, being the SCI value for *UPC* the best one for the 500 instances (median value of the $SCI = 1.88$). This is because, all the individual indicators of this university are better than the corresponding reservation level (values of its achievement

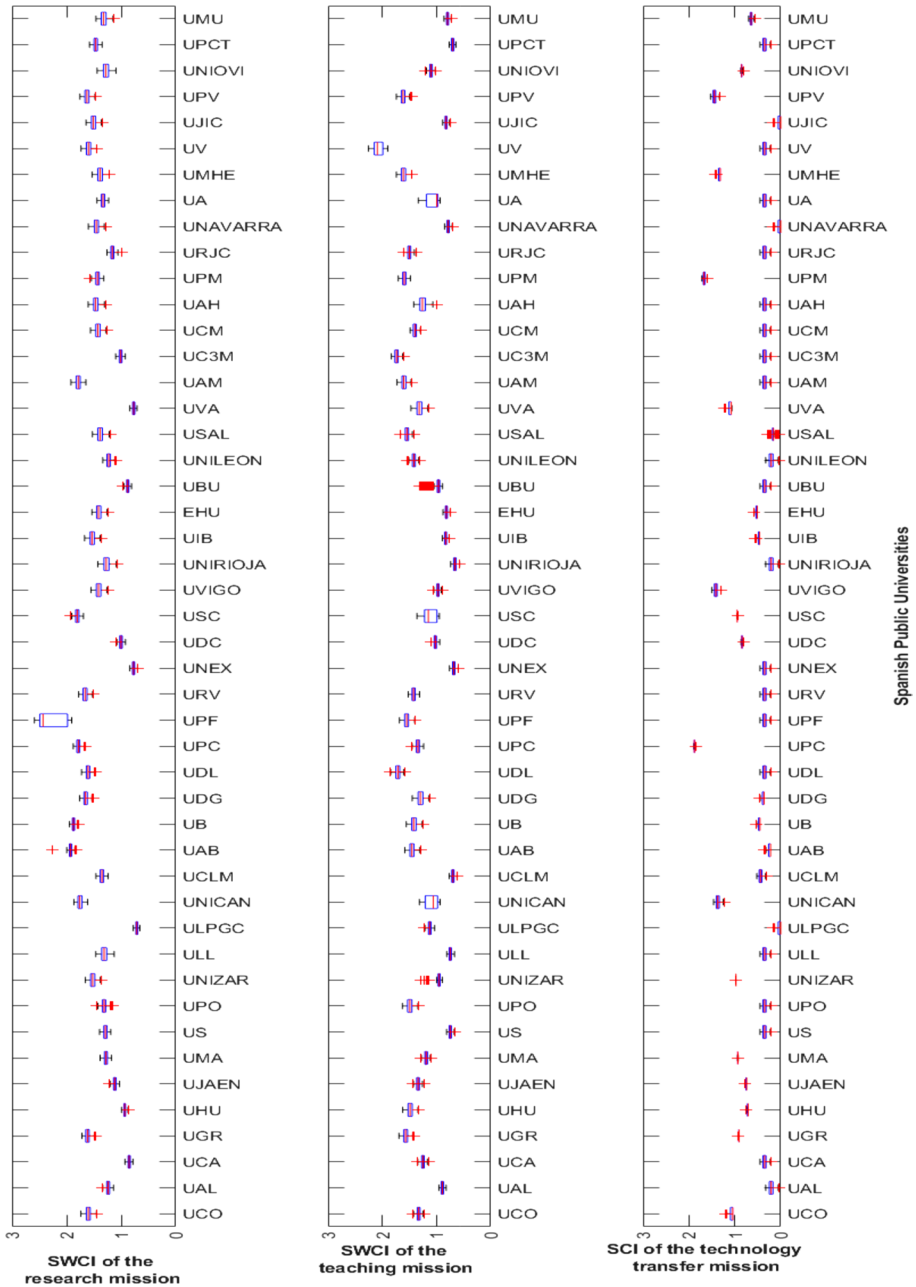


Fig. 4 Robustness analysis on weights: effect on SWCI of each mission

functions > 1). In fact, its worst indicators correspond to the *PCT extensions per doctor* and *Spin-offs per doctor* (1.85 and 1.86, respectively). These two indicators are the lowest

weighted ones in the technology transfer mission (0.23 and 0.19, respectively). This way, its *SCI* is not so bad, since its worse performance corresponds to the lowest weighted indicator.

The rest of universities are in the “poor” performance level. Let us analyze, for instance, the case of *ULPGC*. This university has the worst possible value (0), that is, the worst value among all the Spanish universities, in the highest weighted indicator (*National patents per doctor*, 0.287). This way, for most instances, its *SCI* takes the worst possible value (0).

Regarding the efficiency analysis in the non-compensatory scenario, 17 Spanish public universities are efficient (colored in red in Fig. 5). Once again, the efficiency scores for all the Spanish public universities proved to be stable for the 500 instances, specially for the efficient ones. The differences in the efficiency assignments under the two scenarios are discussed in further detail in “Discussion”.

Comparing Figs. 3 and 5, we can observe that four Spanish public universities (*UAB*, *UAM*, *USAL* and *UJAEN*) are efficient in the compensatory scenario (Fig. 3), while in the non-compensatory one they are inefficient (Fig. 5). First, it can be noticed that the median DEA efficiency scores of *UAB*, *UAM*, *USAL* and *UJAEN* are 1.02, 1.01, 1.01 and 1.07, respectively. This suggests that in response to a small proportional increase (only 2%, 1% and 7%, respectively) of their level of outputs (*SWCI* of each mission, that is, concentrating this increase in the indicators or sub-blocks with the worst behaviors). Note that additional reductions in some inputs and/or increases in some outputs may be necessary.

All these universities perform worse than the reservation level of the technology transfer mission (“poor” performance level). Specifically, in the case of *UAB*, our analysis shows that, despite having one of the best *SWCI* in the research mission, the value of its *SCI* for the technology transfer mission has still much room for improvement. This university performs poorly in the highest weighted technology transfer indicator (median value of its *SCI* in the *National patents per doctor* is 0.22). Similarly, *UAM* has one of the best *SWCI* in the research and teaching missions. However, its *SCI* for the technology transfer mission is in the “poor” performance level. Therefore, these universities should not neglect any of the higher education missions, and pay special attention to the technology transfer mission, since our results reveal that they have a good performance in the teaching and research missions, but they perform poorly the technology transfer one.

On the other hand, *UNICAN* is a medium size university, which is inefficient in the compensatory scenario (Fig. 3), while in Fig. 5, it manages to make an efficient use of its resources (Fig. 5). This university has one of the lowest value of *Faculty members with PhD*, and is among the Spanish public universities with the best *SWCI* performance in the research and technology transfer missions. In fact, as mentioned in the previous analysis, only 8 Spanish public universities are in the “fair” performance level, being *UNICAN* among them (median value of its *SCI* = 1.37 > reservation level = 1).

Summing up, our analysis highlights the convenience of assessing the university performance by using both compensatory scenario (*WCI*) and non-compensatory one (*SCI*), where the *WCI* gives an overall performance measure of the universities, while the *SCI* helps to detect the actions needed to improve the performances of the universities. Furthermore, combining the *MRP-WSCI* composite indicators with the DEA analysis gives us a more complete picture of the institutions under assessment, allowing universities to better understand their weaknesses and strengths.

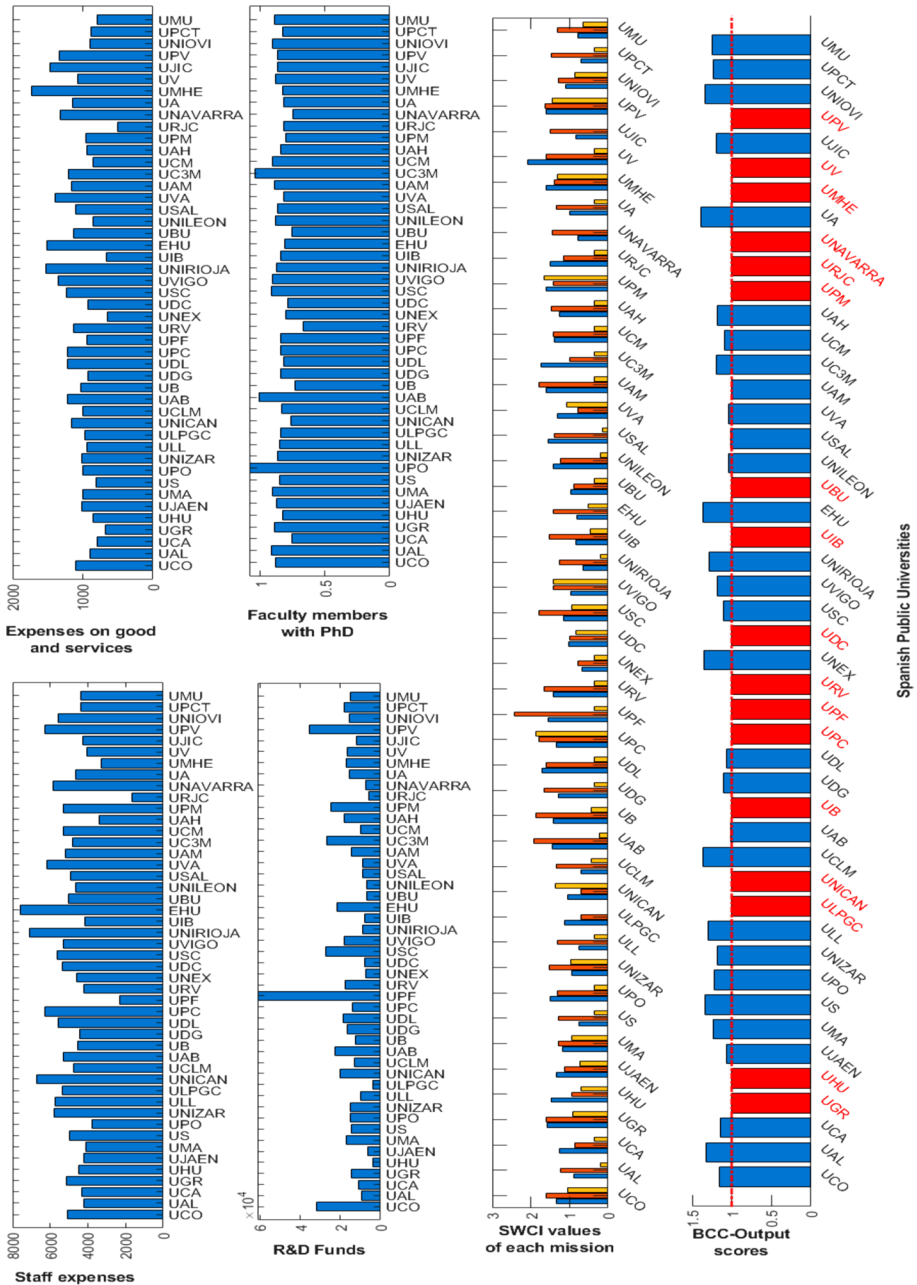


Fig. 5 Inputs, outputs and BCC-output scores of the and SWCI of the three missions

Discussion

In this section, we will discuss certain relevant aspects related methodology proposed for the assessment of universities. First, we report the results of a robustness analysis on the reference levels used. Second, the impact of the weights used is discussed, and results obtained with equal weights are reported and commented. Third, we discuss and compare the results obtained for the compensatory and non-compensatory scenarios. Finally, we give some hints about how this methodology can be used to make improvement decisions for a given university.

Robustness analysis on reference levels

With respect to the *MRP-WSCI* composite indicators, logically, the results obtained depend on the reference levels chosen, which must be set taking into account the aim of the analysis. According to the experts opinions, in this paper, the reference levels have been set by considering the 47 Spanish public universities, and this way, they are compared among themselves. In order to test how a change of the reference levels would affect the *MRP-WSCI* composite indicators, a robustness analysis, with respect to changes of the reference levels has been carried out. To this end, similarly to the case of the weights, we have implemented 500 computation rounds for each option analyzed in “[Application to the Spanish public higher education system](#)”, changing randomly $\pm 10\%$ the Spanish universities reference levels (Fig. 7 in the Appendix shows the results of the 500 instances for both *MRP-WSCI* composite indicators for each mission). First, we can conclude that, in general terms, a high stability can be detected in both *MRP-WSCI* composite indicators for each mission. Second, in general, results with respect to changes of the reference levels are even more robust than results with respect to changes of the weights.

An analysis of the overall results changing the weights

In this paper, given that the purpose is to provide university managers with tools to support their decisions, it has been decided to use weights assigned by experts recognized as such by the decision-making centers. Anyway, as mentioned in “[Introduction](#)”, weights do have an important impact on the overall composite indicator and the results obtained. For this reason, an analysis of the performance and efficiency of the Spanish public university system has been carried out, considering slight modifications (random changes $\pm 10\%$) of the weights given by the experts. In any case, in this section we want to go a step further and discuss the results that would be obtained using a totally different weighting scheme: equal weights for both simple indicators and sub-blocks. In this way, we can clarify the impact of the weights on the solutions obtained.

The comparison of the results obtained can be seen in Fig. 8 of the Appendix. Note that, in the case of equal weights, only one problem has been solved, so the statistical values refer to the values obtained for the different universities. On the other hand, the statistical values of the experts’ weights case refer to the values obtained in the 500 instances, for the different universities. In the compensatory scenario (*WWCI*), it can be seen that the median and mean values of the research and teaching missions improve when considering equal weights, while there is a decrease in the range of values due to a decrease in the maximum values. On the other hand, the results worsen in general in the technology transfer mission, which suggests that the universities present worse behaviors in indicators to which

the experts assigned greater weights. These variations cause greater homogeneity among the universities studied, so the efficiency scores improve. As a result, in the case of equal weights, one more efficient university appear, apart from the 18 that were efficient with the experts' weights.

In the non-compensatory scenario (*SWCI*), the worsening of the performance measures when using equal weights is generalized in the three missions and especially significant in the technology transfer one. This is because a bad behavior in any indicator is regarded as equally bad when building the composite indicator. Once again, the result of this general decline is that efficiency scores improve. As a result, in the case of equal weights, up to 6 more efficient universities appear, apart from the 17 that were efficient with the weights of the experts. Therefore, it can be concluded that the use of equal weights is less demanding when establishing efficient universities. The experts' weights place more demand on certain indicators and sub-blocks, which translates into the existence of less efficient universities.

A comparison of the compensatory and non-compensatory scenarios

As mentioned in “[Combining the MRP-WSCI approach and the DEA analysis](#)”, in the first aggregation step, a full compensation among the individual indicators is allowed, while in the second step, depending on the compensation degree allowed among the different sub-blocks, two different composite indicators are provided (*WWCI* and *SWCI*). On the one hand, the *WWCI* allows to analyse the overall performance of the universities, and on the other hand, the *SWCI* offers alert signs which let the decision-maker detect improvement areas. This way, due to the philosophy behind their construction, there are obvious differences among the results obtained in both scenarios. In order to allow a more direct interpretation of the results obtained, in this section the weights assigned by the experts are used directly, without taking into account the previously described robustness analysis.

Figure 6 shows the levels of efficiency for each university in both scenarios. Each university is represented by a dot, whose horizontal component is its DEA score in the compensatory scenario, and whose vertical component is its DEA score in the non-compensatory scenario. Among the 47 Spanish public universities in the data set, 18 are efficient in both scenarios, and thus, they share the dot placed in (1,1). Their names are colored in red in the figure. There are two other universities placed on the vertical line $x = 1$, which means that they are efficient in the compensatory scenario, but not in the non-compensatory one (colored in green). This means that they perform efficiently in an overall fashion, but some sub-block (or indicator for the technology transfer mission) is performing inefficiently, given the inputs available. Conversely, there are two universities in the horizontal line $y = 1$, colored in blue, which are efficient under the non-compensatory scenario, but they are inefficient in the compensatory scenario. This means that no single sub-block performs in a too poor way given the available inputs, but the overall performance is not good enough to reach efficiency. The rest of the universities are inefficient in both scenarios, with the inefficiency growing as they are placed more to the left and the bottom of the figure. The results show that, on average, the scores among the Spanish public universities are quite similar in both analyses (1.11 and 1.12, respectively). This indicates that, given their available inputs, Spanish universities should expand, on average, their outputs by 11% in the compensatory scenario and 12% in the non-compensatory one, in order to operate efficiently. If we look at the bisector line drawn in Fig. 6, we can see that most of these inefficient universities (17) are below the line, meaning that their inefficiency in the

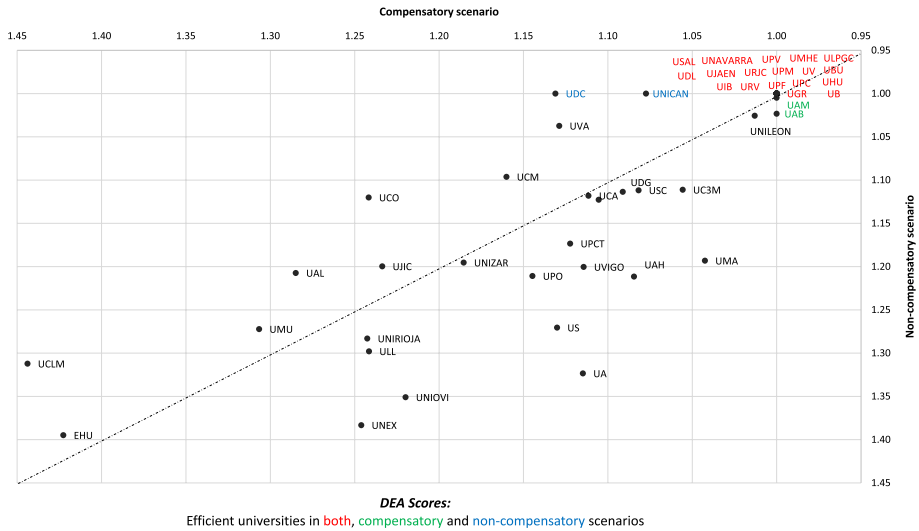


Fig. 6 Efficiency scores for the Spanish public universities in the compensatory and non-compensatory scenarios considering the weights giving by the experts

non-compensatory scenario is greater than that of the compensatory scenario. This most common case happens when at least one sub-block is performing significantly worse than the rest, which gives hints to decision makers about fields of improvement. Finally, there are 8 universities above the line, which means that many sub-blocks or indicators are performing in an inefficient way, although none of them is significantly worse than the rest. In our opinion, these findings show the usefulness of simultaneously considering both scenarios for the DEA efficiency analysis.

Analyzing particular universities: how to move towards efficiency

In what follows, we will use the results obtained by two universities (*UDC* and *UAB*), to illustrate the information provided by combining the *MRP-WSCI* approach and DEA analysis and its usefulness for decision making purposes. As mentioned in “Available system of indicators”, in this paper, two DEA analyses with different outputs are performed, while the inputs are the same in both analyses (*Ratio of faculty members with PhD, R & D funds per faculty member, Staff expenses per students and Expenses on good and services per student*). Our analysis shows that *UAB* is efficient in the compensatory scenario, while *UDC* is efficient in the non-compensatory one.

The achievement functions of the individual indicators, the *WCI* of each sub-block and the *WWCI* and *SWCI* of each mission are provided in Table 3. First, using the *MRP-WSCI* approach, the performance of both universities is analyzed in Table 3a. The scores below 1 appear in red (values worse than the reservation, “poor” performance level); the scores between 1 and 2 are in yellow (values better than the reservation level, but worse than the aspiration level, “fair” performance level); and the scores over 2 are highlighted in green (values better than the aspiration level, “good” performance level). Then, the inputs and

Table 3 The combination of the *MRP-WSCI* and DEA approaches for *UDC* and *UAB*

Mission	MRP-WSCI Composite indicators	UDC	UAB	Sub-block	WCI of each sub-block		Individual indicators	Achievement functions		
					UDC	UAB		UDC	UAB	
The MRP-WSCI approach (a)	Research	WWCI	1.09	2.23	Publications (0.43)	1.00	2.57	Publications per doctor (0.31)	1.14	2.57
					Citations per doctor (0.31)	0.76	2.64			
		Q1 publications per doctor (0.38)	1.08	2.52						
		Other research activities (0.23)	1.24	1.87	International collaboration (0.24)	1.49	2.07			
	SWCI	1.18	1.93	Projects (0.35)	1.11	2.05	Theses defended per doctor (0.16)	1.18	2.68	
							Participation in projects (0.36)	1.24	1.30	
	Teaching	WWCI	1.33	1.63	Results (0.55)	1.14	1.87	Official recognition research (0.24)	1.04	1.97
								Training grants and contracts per doctor (0.23)	0.77	2.07
		SWCI	1.14	1.45	External projection (0.45)	1.56	1.33	National projects assigned per doctor (0.38)	1.59	2.06
								European projects assigned per doctor (0.38)	0.83	2.02
Technology transfer	WCI	1.15	1.04				Ratio of postgraduated student (0.18)	1.61	2.20	
							Success rate (0.22)	1.09	2.04	
The DEA analysis (b)	Analysis performed		Outputs		UDC	UAB	Inputs		UDC	UAB
	The compensatory scenario		WWCI of the research mission	1.09	2.23	Staff expenses per student	5351.25	5320.15		
			WWCI of the teaching mission	1.33	1.63	Expenses on good and services per student	928.00	1222.70		
	The non-compensatory scenario		WCI of the technology transfer mission	1.15	1.04	R&D Funds per faculty member	799.42	22580.46		
			SWCI of the research mission	1.18	1.93	Ratio of faculty members with PhD	0.79	1.01		
			SWCI of the teaching mission	1.14	1.45					
			SCI of the technology transfer mission	0.84	0.22					

outputs levels for both compensatory and non-compensatory scenarios are provided in Table 3b.

As can be observed in Table 3a, *UAB* is one of the Spanish public universities with the lowest value of the *SCI* in the technology transfer mission. Note that this mission has no sub-blocks. Then, the construction of the *MRP-WSCI* is made in one step, where the *SCI* reflects the worst value achieved by the Spanish universities in the individual indicators of the technology transfer mission, relativized by the weight of the indicator. *UAB* performs poorly in the highest weighted technology transfer indicator (value of its *SCI* in the *National patents per doctor* is 0.22).

On the other hand, *UDC* also has its *SCI* value worse than the reservation level in the technology transfer mission. However, its *SCI* value is much better than this of *UAB* (0.84 and 0.22, respectively). Specifically, only 13 Spanish public universities manage to have a better *SCI* value in the above mentioned mission. Besides, *UDC* is performing worse in the indicator *PCT extensions per doctor*, whose weight is less than the *National patents per doctor*.

Furthermore, the worst values achieved by *UDC* in the research and teaching missions are better than the corresponding reservation level but worse than the aspiration level (“fair” performance level). With respect to the research mission, *UAB* has a better *SCI* value than *UDC* (1.93 and 1.18, respectively). This is because, the worst performance of *UAB* takes place in the *other research activities* sub-block, which is the lowest weighted sub-block of the research mission, while in the *publications* and *projects* sub-blocks, *UAB* has a good performance (among all the Spanish public universities, it has the second best performance, just worse than *UPF*).

Regarding the compensatory scenario, *UAB* compensates its unfavorable performance in the *National patents per doctor* with a better value in the rest of technology transfer indicators, and its *WCI* in the technology transfer is in the “fair” performance level (*WCI* = 1.04), although this of *UDC* is still better (*WCI* = 1.15). This is because, as commented before, while *UAB* performs poorly in the highest weighted indicator, *UDC* performs poorly in a less weighted indicator, being its value in the *PCT extensions per doctor* indicator one of the best among all the Spanish public universities. Therefore, *UDC* manages to compensate better than *UAB* in the technology transfer mission. However, *UAB* performs much

Table 4 Improvement simulations for *UDC*

DEA Scores (WWCI)		1.131	1							
		Inefficient	Efficient							
Mission	WWCI Values		Sub-block	WCI of each sub-block		Individual indicators	Achievement functions		Individual indicators	
	Original results	Improvement simulations		Original results	Improvement simulations		Original results	Improvement simulations	Original data	Improvement simulations
Research	1.09	1.24	Publications (0.43)	1.00	1.14	Publications per doctor (0.31)	1.14		0.716	
				(+14.32%)		Citations per doctor (0.31)	0.76		1.305	
			Other research activities (0.23)	1.24	1.39	Q1 publications per doctor (0.38)	1.08	1.46	0.351	0.481
				(+11.52%)		International collaboration (0.24)	1.49		0.438	
						Theses defended per doctor (0.16)	1.18		0.146	
						Participation in projects (0.36)	1.24	1.64	0.433	0.635
				(+13.11%)		Official recognition research (0.24)	1.04		1.109	
Teaching	1.33	1.51	Projects (0.35)	1.11	1.25	Training grants and contracts per doctor (0.23)	0.77		0.010	
				(+12.96%)		National projects assigned per doctor (0.38)	1.59		0.096	
			Results (0.55)	1.14	1.32	European projects assigned per doctor (0.38)	0.83	1.20	0.002	0.004
				(+15.27%)		Ratio of postgraduated student (0.18)	1.61		0.209	
						Success rate (0.22)	1.09		0.835	
						Assessment rate (0.18)	0.85	1.33	0.842	0.891
				(+13.11%)		Graduation rate (0.22)	0.86	1.25	0.703	0.753
External projection (0.45)	1.56	1.74	Drop-out rate (0.18)	1.38		0.188				
	(+11.17%)		International attractiveness (0.43)	1.45	1.86	0.003	0.006			
			National students in an Erasmus program (0.19)	1.25		0.025				
			Foreign students in an Erasmus program (0.23)	1.48		0.032				
			Master degrees in a foreign language (or bilingual) (0.15)	2.38		0.451				
Technology transfer	1.15	1.30			National patents per doctor (0.29)	1.29		0.009		
					R&D contracts and services rendered per doctor (0.29)	1.32		4.263		
					PCT extensions per doctor (0.23)	0.80	1.44	0.002	0.006	
			(+13.11%)		Spin-offs per doctor (0.19)	1.10		0.001		

better than *UDC* in the research and teaching missions (*WWCI* = 2.23 and 1.64, 1.63 and 1.33, respectively). Furthermore, combining these results with the DEA analysis, where as can be seen in Table 3b, despite being the inputs levels of *UAB* higher than this of *UDC*, with the exception of the *Staff expenses per students*, *UDC* makes an efficient use of its resources in the non-compensatory scenario, while in the compensatory one, *UDC* is inefficient, being *UAB* efficient.

Summing up, when the overall performance (*WWCI*) is analyzed, *UAB* is in the “fair” performance level for the teaching and technology transfer missions ($1 < WWCI < 2$), and in the “good” performance level for the research mission ($WWCI > 2$), while *UDC* is in the “fair” performance level for all the three missions. Specifically, the non-compensatory analysis shows that *UAB* has still much room for improvement in the technology transfer mission. Thus, these findings should encourage *UAB* to pay greater attention to the technology transfer mission. Moreover, *UDC* does not perform poorly in any sub-block, since its worst performance is not that bad, but given its input levels, *UDC* should improve its overall performance.

Based on the information obtained in the previous analysis, decision makers can be provided with a tool to simulate improvements in certain indicators and see their impact on the values of the composite indicators and on the efficiency scores. As an example, we will now make a proposal for *UDC*. As previously mentioned, this university is inefficient for the compensatory indicator, with an efficiency score of 1.131. This implies that it must experience an improvement of about 13.1% in the *WWCI* of each of the missions to achieve efficiency (if there is no slack in any of the inputs). Table 4 shows a sensible proposal on how this improvement could be achieved. In the table, the indicators that have been improved in the simulation have been highlighted and for them the improved value is shown next to the current one. In addition, the impact (absolute and in percentage) of these improvements in the composite indicators of the sub-blocks and missions can be seen. The strategy followed has been to improve indicators with worse behavior and those with good behavior and with high weights. Among them, those that are estimated to be most immediately improvable have been chosen. As a result of the improvements in the chosen indicators, a fairly homogeneous improvement has been achieved in all the sub-blocks, and just the desired one in each of the three missions. Obviously, this is a hypothetical proposal and

the real one would have to be implemented by the university authorities, but this simulation shows the potential usefulness of the tool provided by this methodology.

For the case of the *UAB*, which is inefficient in the non-compensatory scenario, the simulation should be based more on improving the indicators and sub-blocks with the worst performance. Finally, a university that is inefficient in both scenarios should combine both strategies to improve its scores. As a conclusion, we can say that the combination of the *MRP-WSCI* composite indicators and DEA analysis provides a rich amount of information. First, the joint consideration of the compensatory and non-compensatory scenarios is highly suitable, where the former gives an overall measure of the universities, while the latter helps to detect the actions needed to improve the performances of the universities. Second, combining both scenarios with the DEA analysis allows universities to better understand their weaknesses and strengths and provides a valuable tool for decision making purposes.

Conclusions

The evaluation of multidimensional frameworks, such as the higher education field, is a complex decision making problem which involves multiple criteria. Within this context, multiple criteria decision making provides a set of tools and techniques that can support such complex decision making processes. In this paper, we propose a combination, in a robust environment, of the *MRP-WSCI* approach with the DEA analysis, where all the aspects regarded as relevant to assess university performance are included as outputs.

Specifically, in order to illustrate the information provided by combining both approaches in such complex decision-making process, we have used the data of 47 Spanish public universities for the academic year 2016–2017. It is important to point out that the main emphasis of this paper is on the methodological contribution, while the evaluation of the Spanish public university system serves as an illustration of the combination proposed. Anyway, in order to detect and improve important factors of the educational efficiency and effectiveness of universities, the approach proposed can be used as a supporting tool for decision making, allowing the decision makers to select appropriate solutions and identify weaknesses and strengths of the universities analyzed, using multiple criteria.

From the practical point of view, our results highlight the need to assess the university performance by using both compensatory and non-compensatory scenarios, where the overall performance measure is given by the former, while the latter allows the decision maker to detect improvement areas. Furthermore, combining the use of composite indicators with the DEA analysis gives us a more complete picture of the universities under assessment, allowing us to consider the inherently multi-faceted academic quality in higher education.

As further research lines, we would like to extend this analysis to a dynamic scheme, in order to study the evolution of the universities along a given period, either by using fixed reference levels for the whole period, or even changing these levels according to new requirements established by the academic authorities. Besides, it would also be interesting to make a wider comparison among universities from different countries, where the results will give an idea of the relative position of each Spanish university with respect to all international universities.

Appendix

Table 5 List of universities analyzed and their abbreviations

University	Abbreviation
U. de Córdoba	UCO
U. de Almería	UAL
U. de Cádiz	UCA
U. de Granada	UGR
U. de Huelva	UHU
U. de Jaén	UJAEN
U. de Málaga	UMA
U. de Sevilla	US
U. Pablo de Olavide	UPO
U. de Zaragoza	UNIZAR
U. de La Laguna	ULL
U. de Las Palmas de Gran Canaria	ULPGC
U. de Cantabria	UNICAN
U. de Castilla-La Mancha	UCLM
U. Autónoma de Barcelona	UAB
U. de Barcelona	UB
U. de Girona	UDG
U. de Lleida	UDL
U. Politècnica de Catalunya	UPC
U. Pompeu Fabra	UPF
U. Rovira i Virgili	URV
U. de Extremadura	UNEX
U. de A Coruña	UDC
U. de Santiago de Compostela	USC
U. de Vigo	UVIGO
U. de la Rioja	UNIRIOJA
U. de les Illes Balears	UIB
U. del País Vasco	EHU
U. de Burgos	UBU
U. de León	UNILEON
U. de Salamanca	USAL
U. de Valladolid	UVA
U. Autónoma de Madrid	UAM
U. Carlos III de Madrid	UC3M
U. Complutense de Madrid	UCM
U. Alcalá de Henares	UAH
U. Politècnica de Madrid	UPM
U. Rey Juan Carlos	URJC
U. Pública de Navarra	UNAVARRA
U. de Alicante	UA
U. Miguel Hernández de Elche	UMHE
U. de Valencia	UV
U. Jaume I de Castellón	UJIC
U. Politècnica de Valencia	UPV

University	Abbreviation
U. de Oviedo	UNIOVI
U. Politécnica de Cartagena	UPCT
U. de Murcia	UMU

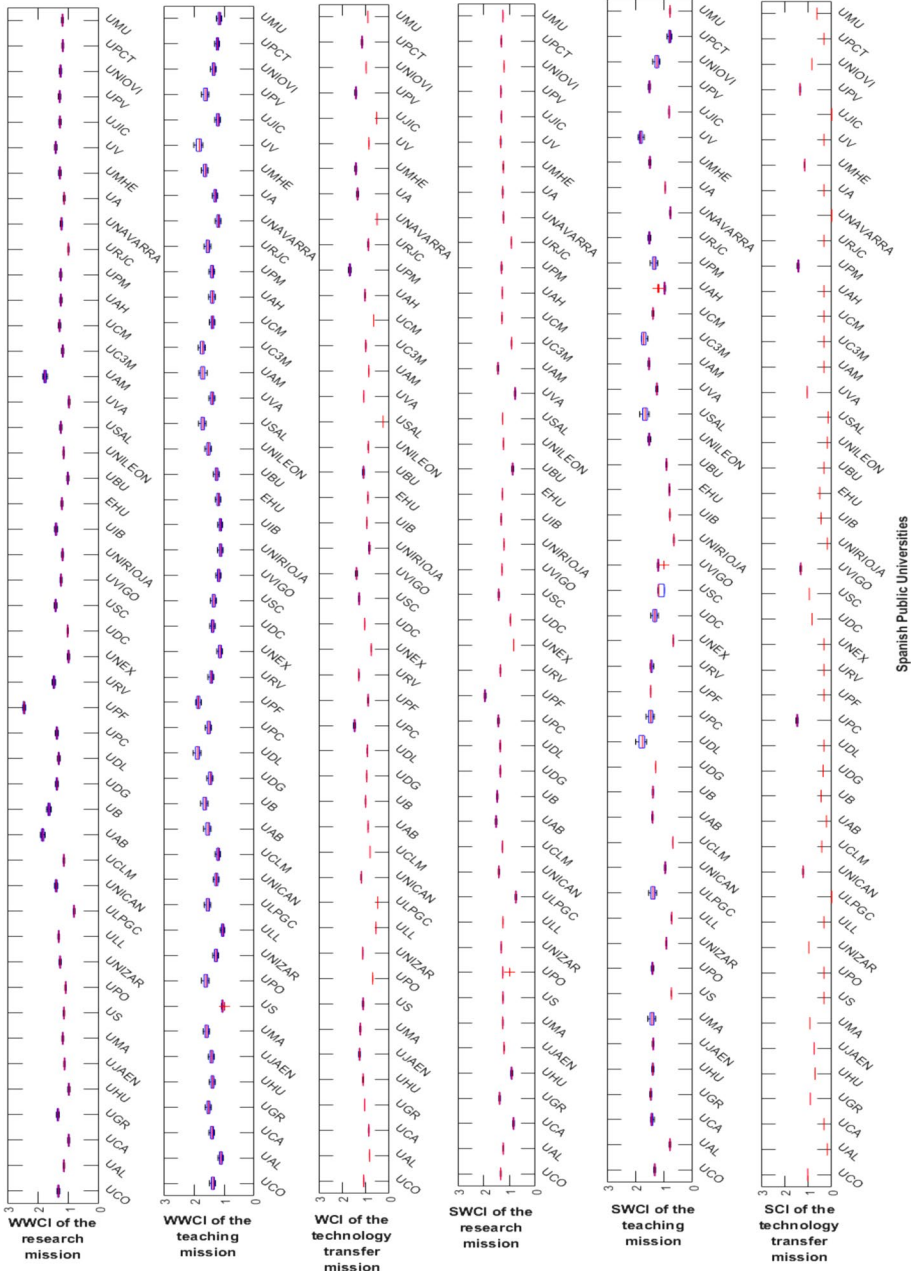


Fig. 7 Robustness analysis on reference levels: effect on MRP-WSCI composite indicators of each mission

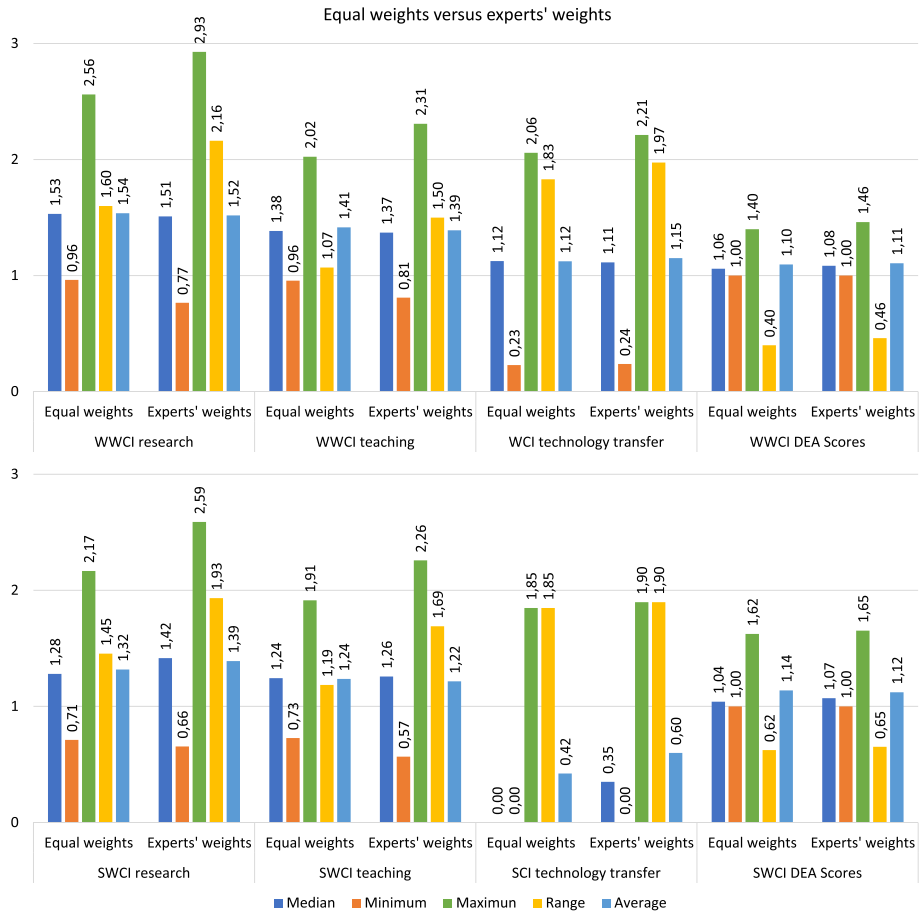


Fig. 8 Comparison of the results by considering equal weights versus experts' weights

Acknowledgements The authors would like to thank the financial support received from the Spanish Ministry of Science and Innovation (Project PID2019-104263RB-C42), from the Regional Government of Andalucía (Project P18-RT-1566), and from the FEDER-UMA (Project UMA18-FEDERJA-065).

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. Funding for open access charge: Universidad de Málaga / CBUA.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abbott, M., & Doucouliagos, C. (2003). The efficiency of Australian universities: A data envelopment analysis. *Economics of Education Review*, 22(1), 89–97.
- Adler, N., & Yazhemsky, E. (2010). Improving discrimination in data envelopment analysis: PCA-DEA or variable reduction. *European Journal of Operational Research*, 202, 273–284.
- Agasisti, T., Barra, C., & Zotti, R. (2019). Research, knowledge transfer, and innovation: The effect of Italian universities' efficiency on local economic development 2006–2012. *Journal of Regional Science*, 59(5), 819–849.
- Agasisti, T., & Pérez-Esparrells, C. (2010). Comparing efficiency in a cross-country perspective: the case of Italian and Spanish state universities. *Higher Education*, 59, 85–103.
- Aguillo, I. F., Ortega, J. L., & Fernández, M. (2008). Webometric ranking of world universities: Introduction, methodology, and future developments. *Higher Education in Europe*, 33(2–3), 233–244.
- Akbari, N., Jones, D., & Treloar, R. (2020). A cross-European efficiency assessment of offshore wind farms: A DEA. *Renewable Energy*, 151, 1186–1195.
- Alasehir, O., Perit-Cakir, M., Acarturk, C., Baykal, N., & Akbulut, U. (2014). URAP-TR: A national ranking for Turkish universities based on academic performance. *Scientometrics*, 101, 159–178.
- Anderson, T. R., Daim, T. U., & Lavoie, F. F. (2007). Measuring the efficiency of university technology transfer. *Technovation*, 27(5), 306–318.
- Antanasijevic, D., Pocajt, V., Ristic, M., & Peric-Grujic, A. (2017). A differential multi-criteria analysis for the assessment of sustainability performance of European countries: Beyond country ranking. *Journal of Cleaner Production*, 165, 213–220.
- Attardi, R., Cerreta, M., Sannicandro, V., & Torre, C. M. (2017). Non-compensatory composite indicators for the evaluation of urban planning policy: The land-use policy efficiency index (LUPEI). *European Journal of Operational Research*, 264(2), 491–507.
- Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, 30(9), 1078–1092.
- Berbegal-Mirabent, J. (2018). The influence of regulatory frameworks on research and knowledge transfer outputs: An efficiency analysis of Spanish public universities. *Journal of Engineering and Technology Management*, 47, 68–80.
- Berbegal-Mirabent, J., Lafuente, E., & Solé, F. (2013). The pursuit of knowledge transfer activities: An efficiency analysis of Spanish universities. *Journal of Business Research*, 66, 2051–2059.
- Blancas, F. J., Caballero, R., González, M., Lozano-Oyola, M., & Pérez, F. (2010). Goal programming synthetic indicators: An application for sustainable tourism in Andalusian coastal counties. *Ecological Economics*, 69, 2158–2172.
- Charles, V., Aparicio, J., & Zhu, J. (2019). The curse of dimensionality of decision-making units: A simple approach to increase the discriminatory power of data envelopment analysis. *European Journal of Operational Research*, 279(3), 929–940.
- Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2, 429–444.
- Charnes, A., Cooper, W. W., & Rhodes, E. (1981). Evaluating program and managerial efficiency: An application of data envelopment analysis to program follow through. *Management Science*, 27(6), 668–697.
- de Jorge Moreno, J., González Robles, A., Martínez, A., Minero Calvo, R., & Georgiana Miron, A. (2019). Assessing efficiency in the Spanish public universities through comparative non-radial and radial data envelopment analysis. *Tertiary Education and Management*, 25, 195–210.
- de la Torre, E. M., Agasisti, T., & Pérez-Esparrells, C. (2017a). The relevance of knowledge transfer for universities' efficiency scores: An empirical approximation on the Spanish public higher education system. *Research Evaluation*, 26(3), 211–229.
- de la Torre, E. M., Gómez-Sancho, J. M., & Pérez-Esparrells, C. (2017b). Comparing university performance by legal status: A Malmquist-type index approach for the case of the Spanish higher education system. *Tertiary Education and Management*, 23(3), 206–221.
- De Witte, K., & López-Torres, L. (2017). Efficiency in education: A review of literature and a way forward. *Journal of the Operational Research Society*, 68, 339–363.
- Dyson, R. G., Allen, R., Camanho, V. V., Sarrico, C. S., & Shale, E. A. (2001). Pitfall and protocols in DEA. *European Journal of Operational Research*, 132(2), 245–259.
- El Gibari, S., Gómez, T., & Ruiz, F. (2018). Evaluating university performance using reference point based composite indicators. *Journal of Informetrics*, 12(4), 1235–1250.
- El Gibari, S., Gómez, T., & Ruiz, F. (2019). Building composite indicators using multicriteria methods: A review. *Journal of Business Economics*, 89, 1–24.

- Expósito-García, A., & Velasco-Morente, F. (2018). How efficient are universities at publishing research? A data envelopment analysis of Spanish state universities. *El profesional de la información, 27*, 1108–1115.
- Farrell, M. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society (Series A), 120*, 253–281.
- Giannoulis, C., & Ishizaka, A. (2010). A web-based decision support system with ELECTRE III for a personalised ranking of British universities. *Decision Support Systems, 48*, 488–497.
- González-Garay, A., Pozo, C., Galán-Martín, A., Brechtelsbauer, C., Chachuat, B., Chadha, D., et al. (2019). Assessing the performance of UK universities in the field of chemical engineering using data envelopment analysis. *Education for Chemical Engineers, 29*, 29–41.
- Heitor, M. V., & Horta, H. (2013). Democratizing higher education and access to science: The Portuguese reform 2006–2010. *Higher Education Policy, 27*(2), 1–19.
- Huang, M. H. (2011). A comparison of three major academic rankings for world universities: From a research evaluation perspective. *Journal of Library and Information Studies, 9*(1), 1–25.
- Johnes, J. (2006). Data envelopment analysis and its application to the measurement of efficiency in higher education. *Economics of Education Review, 25*, 273–288.
- Langhans, S. D., Reichert, P., & Schuwirth, N. (2014). The method matters: A guide for indicator aggregation in ecological assessments. *Ecological Indicators, 45*, 494–507.
- Marginson, S., & Van der Wende, M. (2007). To rank or to be ranked: The impact of global rankings in higher education. *Journal of Studies in International Education, 11*, 306–329.
- Martínez-Campillo, A., & Fernández-Santos, Y. (2020). The impact of the economic crisis on the (in)efficiency of public higher education institutions in Southern Europe: The case of Spanish universities. *Socio-Economic Planning Sciences, 71*, 100771.
- Meng, W., Zhang, D., & Liu, W. (2008). Two-level DEA approaches in research evaluation. *Omega, 36*, 950–957.
- Molinos-Senante, M. (2018). Comparative energy efficiency of wastewater treatment technologies: A synthetic index approach. *Clean Technologies and Environmental Policy, 20*, 1819–1834.
- Murias, P., de Miguel, J. C., & Rodríguez, D. (2008). A composite indicator for university quality assessment: The case of Spanish higher education system. *Social Indicators Research, 89*, 129–146.
- Nardo, M., Saisana, M., Saltelli, A., & Tarantola, S. (2005). *Tools for composite indicators building. Technical report*. Ispra: European Commission.
- Nguyen, B. H., & Zelenyuk, V. (2021). Aggregation of outputs and inputs for DEA analysis of hospital efficiency economics, operations research and data science perspectives. In J. Zhu & V. Charles (Eds.), *Data-enabled analytics* (pp. 123–158). Springer.
- Ruiz, F., Cabello, J. M., & Luque, M. (2011). An application of reference point techniques to the calculation of synthetic sustainability indicators. *Journal of the Operational Research Society, 62*, 189–197.
- Ruiz, F., El Gibari, S., Cabello, J. M., & Gómez, T. (2020). MRP-WSCI: Multiple reference point based weak and strong composite indicators. *Omega, 95*, 102060.
- Salas-Velasco, M. (2020a). Measuring and explaining the production efficiency of Spanish universities using a non-parametric approach and a bootstrapped-truncated regression. *Scientometrics, 122*, 825–846.
- Salas-Velasco, M. (2020b). The technical efficiency performance of the higher education systems based on data envelopment analysis with an illustration for the Spanish case. *Educational Research for Policy and Practice, 19*, 159–180.
- Szuwarzynski, A. (2018). Benefit of the doubt approach to assessing the research performance of Australian universities. *Higher Education Quarterly, 73*(2), 235–250.
- Torres-Salinas, D., Moreno-Torres, J. G., Delgado-López-Cózar, E., & Herrera, F. (2011). A methodology for institution-field ranking based on a bidimensional analysis: The IFQ²A index. *Scientometrics, 88*, 771–786.
- Tukey, W. J. (1977). *Exploratory data analysis*. Addison-Wesley.
- Visbal-Cadavid, D., Martínez-Gómez, M., & Guijarro, F. (2017). Assessing the efficiency of public universities through DEA. A case study. *Sustainability, 9*(8), 1–19.
- Wierzbicki, A. P. (1980). The use of reference objectives in multiobjective optimization. In G. Fandel & T. Gal (Eds.), *Lecture notes in economics and mathematical systems* (Vol. 177, pp. 468–486). Springer-Verlag.
- Wierzbicki, A. P., Makowski, M., & Wessels, J. (Eds.). (2000). *Model-based decision support methodology with environmental applications*. Kluwer Academic Publishers.
- Yang, G. L., Fukuyama, H., & Song, Y. Y. (2018). Measuring the inefficiency of Chinese research universities based on a two-stage network DEA model. *Journal of Informetrics, 12*, 10.

- Zhong, W., Yuan, W., Li, S. X., & Huang, Z. (2011). The performance evaluation of regional R &D investments in China: An application of DEA based on the first official China economic census data. *OMEGA*, *39*(4), 447–455.
- Zhou, P., Fan, L. W., & Zhou, D. Q. (2010). Data aggregation in constructing composite indicators: A perspective of information loss. *Expert Systems with Applications*, *37*, 360–365.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.