Decisions in Economics and Finance (2020) 43:413–441
https://doi.org/10.1007/s10203-020-00298-4

A three-system approach that integrates DEA, BSC, and AHP for museum evaluation

Antonella Basso[1] · **Stefania Funari**[2]

Received: 7 January 2020 / Accepted: 20 July 2020 / Published online: 22 September 2020
© The Author(s) 2020

Abstract
This paper proposes to measure the museums performance with a model that combines the Data Envelopment Analysis (DEA) and Balanced Scorecard (BSC) methodologies with a third method, the analytic hierarchy process (AHP), which is often used to support decision making. Starting from the two-stage DEA–BSC model of Basso et al. (Omega Int J Manag Sci 81:67–84, 2018), which integrates DEA and BSC, we explore the advantages to consider also the AHP methodology, with the aim to include the judgement of some museums' experts on the relative importance of the BSC perspectives in the performance evaluation model. A first approach uses directly the AHP priorities derived from the judgements expressed by the museums' experts interviewed to determine the weights to aggregate the four BSC performance scores into an overall performance indicator. A second approach uses the judgments of the museums' experts indirectly to introduce proper restrictions on the output weights of the second-stage DEA model. With this approach, we overcome the problem arising from the dispersion of the preferences within the group of experts, that may heavily affects the first approach. Both approaches proposed in this contribution are applied to the case study of the municipal museums of Venice.

Keywords Performance evaluation · Data Envelopment Analysis (DEA) · Balanced Scorecard (BSC) · Analytic hierarchy process (AHP) · Group decision making · Museums

JEL Classification C61 · Z11 · C44

✉ Antonella Basso
basso@unive.it

Stefania Funari
funari@unive.it

1 Department of Economics, Ca' Foscari University of Venice, Cannaregio 873, 30121 Venice, Italy

2 Department of Management, Ca' Foscari University of Venice, Cannaregio 873, 30121 Venice, Italy

1 Introduction

The importance of measuring the performance of museums is widely highlighted in the literature and may have different motivations (see for example Basso and Funari 2004). One reason is related to the management of museums resources, generally scarse, which have to be optimized by taking into account the cultural objectives to be pursued. Another reason is related to fundraising; the survival of a cultural organization is often guaranteed by the presence of both private and public funding and this calls for transparency and evidence of an efficient use of the resources.

The present contribution aims at measuring the museum performance by proposing a model that combines the Data Envelopment Analysis (DEA) and Balanced Scorecard (BSC) methodologies with a third methodological tool which is often used to support the decision-making process, namely the analytic hierarchy process (AHP).

Recently, the advantages of integrating DEA with BSC for the evaluation of museum performance have been explored by Basso et al. (2018) with the proposal of a DEA–BSC model which employs a two-stage evaluation process. Starting from such a model, in the present paper we explore the possibility to exploit also the potential of the AHP method in the second stage of the evaluation process, with the aim to include some experts' judgements on the relative importance of the four BSC perspectives in the model.

A first approach is to use directly the weights obtained using the pairwise comparison matrices computed from the judgements of the museums' experts interviewed (the so-called AHP priorities) in order to aggregate the four BSC performance scores computed at the first stage into an overall performance indicator. This direct approach often presents some difficulties linked to the presence of inconsistency in some judgements and/or to an excessive dispersion of preferences of the different experts.

A second approach utilizes the judgments of the museums' experts indirectly, to impose proper restrictions on the weights of the outputs of a second-stage DEA model. This second approach allows us to overcome the problems arising from the differences in the judgements of the various decision makers, since it can be applied in any case, no matter how big the differences.

Both approaches proposed in this contribution are applied to a case study, represented by the set of municipal museums of Venice first analyzed in Basso et al. (2018).

The paper is organized as follows. We first summarize the essential features of the BSC and DEA methodologies and their applications to the evaluation of the performance of museums (Sect. 2). In Sect. 3, we briefly present the two-stage DEA–BSC model developed in Basso et al. (2018) to assess the museums performance. Sect. 4 discusses how the AHP methodology can be used to directly compute the weights for aggregating the BSC performance scores of museums, while Sect. 5 reports the results of the application carried out to the municipal museums of Venice. Section 6 proposes the DEA–BSC–AHP three-system model which uses AHP to define the weights restrictions in the second-stage DEA model. Finally, this model is applied to the municipal museums of Venice in Sect. 7.

2 BSC and DEA in the museum sector

In the literature, we find a few papers that adopts either Balance Scorecard (BSC) or Data Envelopment Analysis (DEA) to assess the performance of museum institutions, while Basso et al. (2018) uses these two methodologies jointly. The following Subsections 2.1 and 2.2 briefly recall the two approaches and mention the main contributions regarding their application to the museum sector, whereas the contribution of Basso et al. (2018) is summarized in Sect. 3.

2.1 Balance Scorecard for museums

One of the main features of Balanced Scorecard (BSC) as a tool to measure firms' performance, is to allow us to evaluate an organization from multiple points of view; in particular, the emphasis is on the balance among the following four perspectives:

1. Economic-Financial (or "Financial") Perspective: This perspective reflects the profitability and the ability to meet the needs of shareholders;
2. Customer Perspective: It concerns the way in which the company should be viewed by customers to be able to implement its vision;
3. Internal processes Perspective: In this perspective, the company makes explicit what must be done internally in an excellent way to meet the expectations of shareholders and customers;
4. Innovation and learning Perspective: It reflects the ability of the company to improve, innovate and learn.

Originally designed for private companies, BSC was subsequently adapted to take into account the peculiarity of nonprofit enterprises which have some social missions besides (or instead of) financial objectives. In this case, the literature suggests to move the financial perspective from the top to the bottom line of the hierarchy, since the financial objective is not the primary aim of nonprofit organizations (Kaplan and Norton 2001).

Within this framework, BSC has also been used to assess the performance of cultural organizations. With regard to this, several advantages have been highlighted, such as the possibility to clarify the museum mission, to translate its strategy into operational goals, to focus just on few key indicators that are related to the critical success factors.

Among the main BSC applications to museum institutions we find: Haldma and Lääts (2011), which presents an empirical case study on a central museum in Estonia, Marcon (2004), that discusses a framework useful to adapt BSC to museums considered as nonprofit organizations, Wei et al. (2008), which describes a disclosure index based on BSC for museums in New Zealand and UK, and Zorloni (2012), that discusses the use of a theoretical BSC framework for the strategic goals and activities of museums and focuses on a set of critical success factors for visual art museums.

These studies propose interesting conceptual schemes within the BSC context but only rarely they lead to concrete applications to cultural organizations. Generally they provide a theoretical definition of sets of desirable performance indicators which altogether can measure different aspects of the museum performance with respect to

each BSC perspective. However, the experiences cited reveal difficulties that often prevent the computation of the metrics in concrete applications.

In general, despite the many strengths of BSC as a strategic management tool, only few concrete applications to museums have been actually carried out. Indeed, while BSC is especially useful in translating an organization's strategic objectives into strategic indicators that are related to the different BSC dimensions, its conceptual framework, often very elaborated, may be difficult to put into practice.

2.2 Data envelopment analysis for museums

As we have mentioned in Sect. 2.1, BSC is mainly used to give a comprehensive vision of a particular organization, rather than to compare the performance of different institutions. With regard to this issue, a useful approach that provides an operational tool which allows to compare the performance of a set of museums is Data Envelopment Analysis.

Indeed, DEA can compute an optimal efficiency indicator for each museum that takes into account several different dimensions that are considered relevant for the museum performance. This efficiency indicator, called efficiency score, is calculated as the ratio between a weighted sum of outputs and a weighted sum of inputs, where the outputs usually represent the products or services provided by the organizations whereas the inputs refer to the resources used to obtain the outputs. The DEA efficiency score allows to synthesize all the dimensions included in the analysis into a single performance indicator that comes from a comparative analysis of the set of museums investigated.

Naturally, different weighting schemes may lead to different values of the efficiency indicator and to different evaluations of the museum's efficiency. The DEA technique overcomes the difficulty to seek a common set of input and output weights by suggesting to choose for each museum the most favorable weights, solving an optimization problem for each of the museums investigated.

Suppose we have n museums to be evaluated and let us consider t outputs and m inputs. Let y_{rj} and x_{ij} be the amount of output r ($r = 1, \ldots, t$) and input i ($i = 1, \ldots, m$) associated with museum j ($j = 1, 2, ..., n$), respectively. Moreover, let us denote by u_r and v_i the weights associated with output r and input i, respectively.

The DEA efficiency scores are computed by solving an optimization problem for each museum. In the basic constant returns-to-scale DEA model, the efficiency score of the target museum o (with $o = 1, 2, ..., n$) is computed by solving the following fractional programming problem:

$$\text{maximize} \quad \frac{\sum_{r=1}^{t} u_r y_{ro}}{\sum_{i=1}^{m} v_i x_{io}} \tag{1}$$

subject to:

$$\frac{\sum_{r=1}^{t} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \leq 1 \qquad j = 1, 2, \ldots, n \tag{2}$$

$$u_r \geq 0 \qquad r = 1, 2, \ldots, t \tag{3}$$

$$v_i \geq 0 \qquad i = 1, 2, \ldots, m \tag{4}$$

As we may see, the input and output weights are determined so as to maximize the efficiency score of museum o, subject to the constraint that the efficiencies of all the museums investigated are less than or equal to one. Therefore the efficient museums have an optimal score equal to 1, which is the maximum value of the efficiency score. Moreover, the optimal weights which solve problem (1)–(4) are computed as the most favorable weights for museum o; thus they will be different for each museum.

Model (1)–(4) is the DEA model most often used in the applications to cultural organizations. However, there are a number of different, more complicated, DEA models that could be used, for a detailed presentation of several DEA models; see, e.g., Cooper et al. (2000).

Actually, different types of models based on the DEA approach have been proposed in the literature to measure the performance of museums. In particular, Pignataro (2002) applies both a CCR (Charnes–Cooper–Rhodes) model and a BCC (Banker–Charnes–Cooper) model as well as the Malmquist index to a set of Sicilian museums; Basso and Funari (2004) adopts both a CCR model and an input-oriented BCC model, carries out a DEA cross-efficiency evaluation and undertakes an empirical analysis on a set of Italian museums; Basso and Funari (2003) compares the results obtained with a CCR and a BCC model with those provided by a FDH (free disposal hull) model on a set of Italian museums; Mairesse and Vanden (2002) analyzes a group of museums from the French-speaking region of Belgium with the following FDH models: FDH-CRS (constant returns-to-scale), FDH-NIRS (non increasing returns-to-scale), FDH-NDRS (non-decreasing returns-to-scale), FDH-RRS (restrictive concept of returns-to-scale); Del Barrio et al. (2009) explores the efficiency of a sample of museums in Spain with principal components and cluster analysis, and CCR, input-oriented BCC and superefficiency models; Del Barrio and Herrero (2014) presents an application to a regional system of Spanish museums obtained with a CCR and an input-oriented BCC model and the Malmquist index; Carvalho et al. (2014) applies a BCC model to a set of Portuguese museums; Taheri and Ansari (2013) examines a set of museums in Tehran with a CCR model and a MAJ (Mehrabian, Alirezaee, Jahanshahloo) full ranking method; Basso et al. (2018) adopts a more general two-stage model that combines the BSC and DEA methodologies and allows for variable returns-to-scale and weights restrictions; Del Barrio-Tellado and Herrero-Prieto (2019) compares the results of a two-stage DEA SBM network model to those obtained with a pure DEA SBM model with and without superefficiency on a set of Spanish state-run museums.

3 A two-stage DEA–BSC model for museum evaluation

The joint adoption of DEA and BSC, two methods coming from different fields, may enhance the process of evaluating the performance of organizations. As mentioned in Sect. 2, the synergies between the two methodologies for the performance evaluation

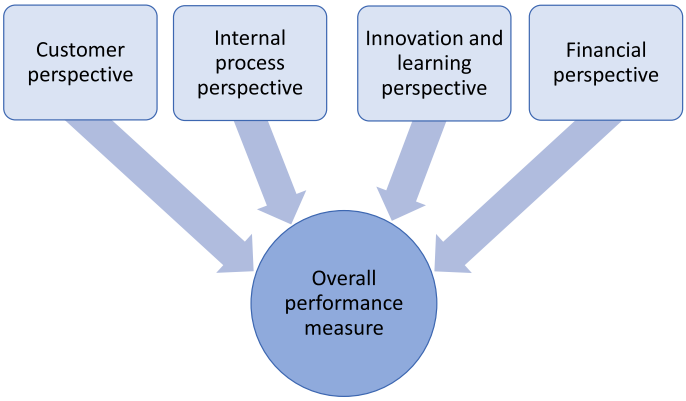

Fig. 1 Basic flowchart of the two-stage DEA–BSC model (source: reworking from Basso et al. (2018))

of museums have been exploited in Basso et al. (2018) with the proposal of a two-stage DEA–BSC model for the evaluation of the performance of a set of museums.

The idea exploited in Basso et al. (2018) is first (in stage 1) to measure the museums performance according to each BSC perspective; a proper DEA model is therefore designed for each BSC perspective, by choosing the input and output variables relevant for that perspective.

As for the connection among the different BSC perspectives, in the museum sector and more in general in the whole cultural sector, we do not find a hierarchy between the BSC perspectives which is clearly defined and generally accepted. This is due to the very nature of the nonprofit organizations that are common in this sector and to the difficulty to measure the fulfilment of the social purpose of their activity, which is directed toward stakeholders with different objectives; for a comprehensive discussion on this issue, see Basso et al. (2018). As a consequence, we have opted for a simultaneous consideration of the different BSC perspectives and therefore the four DEA–BSC models are all set at the same level in the first stage.

At the end of the first stage, four efficiency indicators are therefore obtained, one for each BSC perspective; these indicators capture the different dimensions of the museum's performance. Subsequently, the second stage of the evaluation process uses the DEA approach once again to aggregate the performance scores of the four BSC perspectives into an overall performance indicator. A graphical representation of the basic flow chart of the two-stage DEA–BSC model is given in Fig. 1.

3.1 First stage

The four DEA models associated with the four BSC perspectives in the first stage have a similar structure: they are all variable returns-to-scale DEA models with output orientation and virtual weights restrictions (see Basso et al. 2018).

Table 1 Input and output variables of the four DEA–BSC models of the first stage

Input	Outputs
Customer Perspective model	
Insured value	1. Visitors
	2. Web site visits
	3. Members
	4. Donations
	5. Catalogues
Internal process perspective model	
Total costs	1. Conservation, restoration costs
	2. Amount spent for new acquisitions
	3. Visitors
Innovation and learning perspective model	
Constant	1. Personnel training
	2. Sustainability indicators:
	(2a. Innovative lighting)
	(2b: Environmental sustainability)
	(2c: Facilities for people with disabilities)
Financial perspective model	
Expenditure	1. Income from tickets
	2. Sponsorships, donations, public funding
	3. Other incomes

The input and output variables were carefully selected by designing a special BSC scheme tailored to meet museums' peculiarities. Indeed, the variables of the DEA models focus on the main characteristics of the museum activities which affect each BSC perspective; moreover, the variables have to be computable for all the museums investigated. Table 1 displays the input and output variables of the four DEA–BSC models used at the first stage.

With reference to the DEA model associated with the Customer perspective, the outputs are somewhat related to the stakeholders of the museums; among the output variables we find the number of: museum visitors, Web site visitors, catalogues sold in the museum bookshop, museum members, and the donations received.

The outputs considered in the DEA model for the internal process perspective are instead related to the main purposes of the museum in terms of preservation of cultural heritage (conservation and restoration costs), expansion of museum's collections (amount spent for new acquisitions), increase in value for the museum's customers (measured by the number of visitors).

The personnel training (number of hours per employee) and one or more sustainability indicators make up the outputs of the DEA model for the Innovation and learning perspective. As for the sustainability indicators, we list a few indicators: the presence

of innovative lighting, an indicator of environmental sustainability that measures the adoption of appropriate environmental saving measures, and the number of different facilities for people with disabilities; in the empirical application, we consider a comprehensive indicator computed as a linear combination of such indicators; for the details, see Basso et al. (2018).

Finally, in the financial perspective, the outputs are the different entries that form museum's income (income from tickets, sponsorships, donations, public funding, and other incomes).

As for the inputs of the first-stage DEA–BSC models, each model has a single input: the insured value, as a proxy for the value of the museum exhibits in the customer perspective model; the total operating costs in the internal process perspective model, including the personnel costs that represent the highest share of costs but have not been considered separately in accordance with the suggestions of the BSC approach (other DEA contributions for museums, not using a BSC approach, make a different choice and explicitly consider the personnel costs or number; see, for example, Mairesse and Vanden 2002; Pignataro 2002; Del Barrio-Tellado and Herrero-Prieto 2019); the expenditure (the algebraic sum of operating costs, financial costs and taxes) in the financial perspective model.

A separate consideration has to be made for the innovation perspective model, for which indeed a specific input is not required, since the outputs are a ratio (the number of hours per employee) and a quality indicator (the sustainability indicator). On the other hand, it is possible to show (Lovell and Pastor 1999) that an output oriented DEA model without inputs is equivalent to an output oriented DEA model with a single constant input. In a way, a constant input can be seen as a "dummy input" having the same value for all decision-making units (Cherchye et al. 2007).

As regards the structure of the four DEA models associated with the four BSC perspectives, they all have the same structure: each of them is a variable returns-to-scale DEA model, with output orientation and restrictions on the output weights (see Basso, Casarin, Funari, 2018). The choice of variable returns-to-scale gives more flexibility, since it does not impose that the returns-to-scale are constant, but at the same time, it does not prevent them from being constant.

Actually, we may observe that the innovation perspective model could equivalently be written as a constant returns-to-scale model, since Lovell and Pastor (1999) proved that an output oriented variable returns-to-scale model with a single constant input is equivalent to the corresponding constant returns-to-scale model.

As for the weights restrictions imposed on the DEA models, they serve to ensure that we do not completely overlook some important variables in the computation of museum's efficiency score.

The virtual weights restrictions considered are applied only to the weights of the outputs and set lower and upper limits on the proportion of each virtual output (defined as the product of the level of output r' and the related optimal weight) with respect to the total virtual output:

$$L_{r'} \leq \frac{u_{r'} y_{r'j}}{\sum_{r=1}^{t} u_r y_{rj}} \leq U_{r'} \quad r' = 1, \ldots, t; \quad j = 1, \ldots, n \tag{5}$$

where $L_{r'}$ and $U_{r'}$ represent the lower and upper percentage values set for the virtual output r' (on virtual weights restrictions see (Wong and Beasley 1990).

For example, in the application of the two-stage DEA–BSC model carried out in Basso et al. (2018) on the municipal museums of Venice (the "MUVE museums"), the lower and upper bounds referred to the first output variable "Visitors" in the DEA model designed for the Customer perspective are $L_1^C = 0.4$ and $U_1^C = 0.9$, respectively. These bounds entail that the visitors variable will affect the performance score of the customer perspective model in a percentage that may range from 40 % to 90 %: Since exhibiting its collections to the public is an important museum's mission, we need to give this variable a relevant role in the computation of the performance score. We refer to Table 6 in Basso et al. (2018), p. 79, for the values chosen for all the virtual weights restrictions in the analysis of the MUVE museums.

The "MUVE museums" are a set of 11 municipal museums of Venice managed by the MUVE Foundation, including several renowned museums such as the Doge's Palace and the Correr Museum. We have carried out some empirical investigations applying the DEA–BSC–AHP models on data of the MUVE museums referred to the year 2013. The data have been kindly provided by the MUVE Foundation and are reported in Tables 2, 3, 4 and 5, which show the input and output variables of all DEA–BSC models.

In the analysis carried out with the DEA–BSC–AHP three-system model that will be presented later, in Sects. 5 and 7, the four DEA–BSC models of the first stage coincide with those presented in Basso et al. (2018); of course, since both the data used and the weights restrictions adopted are the same, the results of the first stage coincide with those obtained in Basso et al. (2018) and are reported in Table 6, which summarizes the values of the performance scores obtained for each BSC perspective. It can be seen that the performance of MUVE museums differ according to the BSC perspective considered; moreover, none of the museums is efficient with respect to all perspectives.

3.2 Second stage

The computation of the performance scores related to the different BSC perspectives allows to identify the areas where a museum is stronger or weaker, in comparison with the other museums under evaluation.

However, it is useful to summarize the performances achieved by the museums across the different areas in a synthetic indicator. In a way, the idea is similar to that of the graduation grade, which synthesizes the grades obtained in all exams in a unique score.

A DEA approach, looking for the most favorable weights for each organization, can be exploited also at the second stage, and a synthetic index can be computed starting from the four performance scores obtained at the end of the first stage.

The DEA model developed by Basso et al. (2018) in the second stage has four output variables that coincide with the performance scores associated with the four BSC perspectives, and a constant input (see Table 7). It is once again a variable returns-

Table 2 Values of the input and output variables of the Customer Perspective model for the MUVE museums (percentage values)

Museums	Insured value	Visitors	Web site visits	Members	Donations	Catalogues
Doge's Palace	57.73	65.64	50.80	20.22	98.03	18.44
Correr Museum	12.08	11.73	12.45	33.71	0.00	1.78
Glass Museum	5.13	6.61	5.39	1.12	1.08	39.25
Ca' Rezzonico	7.62	4.95	5.48	12.13	0.88	4.34
Ca' Pesaro	4.67	3.03	5.89	8.31	0.00	3.65
Natural History Museum	4.88	3.11	5.62	11.24	0.00	18.24
Mocenigo Palace	1.95	0.41	3.52	6.97	0.00	0.20
Carlo Goldoni's House	0.82	0.86	1.46	5.17	0.00	3.16
Clock Tower	0.47	0.40	1.33	0.00	0.00	0.00
Lace Museum	0.96	1.29	1.07	0.00	0.00	8.88
Fortuny Palace	3.70	1.96	7.00	1.12	0.00	2.07
Total	100.00	100.00	100.00	100.00	100.00	100.00

Table 3 Values of the input and output variables of the internal process perspective model for the MUVE museums (percentage values)

Museums	Total costs	Conservation, restoration costs	Visitors
Doge's Palace	30.95	8.59	65.64
Correr Museum	19.31	26.73	11.73
Glass Museum	4.01	1.44	6.61
Ca' Rezzonico	10.70	23.53	4.95
Ca' Pesaro	10.70	14.72	3.03
Natural History Museum	9.06	0.59	3.11
Mocenigo Palace	4.34	0.00	0.41
Carlo Goldoni's House	2.77	2.46	0.86
Clock Tower	0.55	18.74	0.40
Lace Museum	1.69	0.00	1.29
Fortuny Palace	5.91	3.21	1.96
Total	100.00	100.00	100.00

Table 4 Values of the input and output variables of the innovation perspective model for the MUVE museums (percentage values)

Museums	Constant	Personnel training	Sustainability indicator
Doge's Palace	1	9.96	10.23
Correr Museum	1	11.48	10.23
Glass Museum	1	8.78	10.23
Ca' Rezzonico	1	9.48	10.23
Ca' Pesaro	1	10.45	10.23
Natural History Museum	1	7.44	10.23
Mocenigo Palace	1	8.44	5.68
Carlo Goldoni's House	1	9.60	10.23
Clock Tower	1	9.67	4.55
Lace Museum	1	7.47	9.09
Fortuny Palace	1	7.24	9.09
Total		100.00	100.00

to-scale DEA model with output orientation and virtual weights restrictions, with a structure similar to the DEA models of the first stage. Of course, this model, too, could equivalently be written as a constant returns-to-scale model given that it has a single constant input.

Table 5 Values of the input and output variables of the financial perspective model for the MUVE museums (percentage values)

Museums	Expenditure	Income from tickets	Sponsorships, donations, public funding	Other incomes
Doge's Palace	30.95	79.11	22.23	78.22
Correr Museum	19.31	7.01	24.80	11.32
Glass Museum	4.01	4.33	0.90	1.19
Ca' Rezzonico	10.70	3.03	2.14	2.73
Ca' Pesaro	10.71	1.91	1.99	1.75
Natural History Museum	9.06	1.37	11.74	1.93
Mocenigo Palace	4.34	0.19	30.42	0.63
Carlo Goldoni's House	2.77	0.36	0.00	0.13
Clock Tower	0.55	0.38	4.87	0.13
Lace Museum	1.69	0.54	0.92	0.15
Fortuny Palace	5.92	1.77	0.00	1.81
Total	100.00	100.00	100.00	100.00

Table 6 Performance scores obtained for each BSC perspective at the first stage of the DEA–BSC model in the assessment of the MUVE museums (reworking from Basso et al. (2018))

Museums	Customer perspective	Internal process perspective	Innovation and learning perspective	Financial perspective
Doge's Palace	1.0000	1.0000	0.9472	1.0000
Correr Museum	0.9064	0.4952	1.0000	0.1711
Glass Museum	1.0000	0.8426	0.9062	0.3695
Ca' Rezzonico	0.5797	0.5120	0.9305	0.1110
Ca' Pesaro	0.5959	0.3180	0.9642	0.0702
Natural History Museum	0.6136	0.1668	0.8600	0.0691
Mocenigo Palace	0.3589	0.0476	0.6775	0.0334
Carlo Goldoni's House	1.0000	0.1666	0.9348	0.0493
Clock Tower	1.0000	1.0000	0.7145	1.0000
Lace Museum	1.0000	0.4512	0.7939	0.1321
Fortuny Palace	0.5570	0.1736	0.7860	0.1256

Table 7 Input and output variables of the DEA–BSC model at the second stage

Input	Outputs
Constant	1. Customer perspective score
	2. Internal process perspective score
	3. Innovation and learning perspective score
	4. Financial perspective score

In the empirical application to the MUVE museums, the following lower and upper bounds were chosen to restrict the virtual output weights of the DEA model in the second stage:

Customer performance score	$L_1^2 = 0.2$	$U_1^2 = 0.8$
Internal process performance score	$L_2^2 = 0.05$	$U_2^2 = 0.5$
Innovation and learning performance score	$L_3^2 = 0.05$	$U_3^2 = 0.5$
Financial performance score	$L_4^2 = 0.05$	$U_4^2 = 0.5$

These restrictions emphasize the importance of the Customer perspective in the museum context but ensure at the same time that no BSC perspective is overlooked in the computation of the overall performance indicator. Let us point out that, though in the DEA–BSC model at the second stage the four BSC perspectives are set all at the same level, without imposing a hierarchical order, the use of virtual weights restrictions allows one to somewhat grade their importance.

The values of the overall performance indicator computed for the MUVE museums in the second stage of the evaluation process are reported in Table 8, together with the relative position in the ranking.

4 Using AHP to weight the performance scores of museums

We have seen that the use of a DEA approach to aggregate the four BSC performance scores into an overall performance indicator leads to a different weighting scheme for each museum, as for each organization the most favorable set of weights are selected.

On the one hand, this flexibility in the determination of the weights can be viewed as a special advantage of DEA; on the other hand, sometimes the museum's decision

Table 8 Overall indicator and ranking obtained at the second stage with the DEA–BSC model for the MUVE museums (reworking from Basso et al. (2018))

Museums	Overall indicator	Ranking position
Doge's Palace	1.0000	1
Correr Museum	0.5283	4
Glass Museum	0.7155	3
Ca' Rezzonico	0.4278	6
Ca' Pesaro	0.3743	8
Natural History Museum	0.3293	10
Mocenigo Palace	0.2053	11
Carlo Goldoni's House	0.4276	7
Clock Tower	0.9840	2
Lace Museum	0.5081	5
Fortuny Palace	0.3311	9

makers may wish to compare the performance of cultural institutions according to a common weighting structure reflecting their value judgments.

Accordingly, if the preferences of the decision makers were known, they could be taken into account in the construction of the overall performance indicator. To this aim, a simple weighted scoring approach could be adopted in the second stage of the evaluation process in place of a DEA model. Indeed, if we already knew the weights associated with each BSC perspective, we might compute an overall performance indicator as the weighted average of the performance scores of the different BSC perspectives. As a limiting case, by assigning equal importance, i.e., equal weights, to all perspectives we would simply obtain the arithmetic mean.

However, it is well known that it is often hard, if not even impracticable, to obtain from the museums' decision makers spontaneous indications on the values of the weights. In this context, the widely known analytic hierarchy process (AHP) methodology could be useful as it provides an operational tool, based on the elicitation of decision makers' preferences, which leads to the determination of the weights associated with the elements of a decision making problem.

Essentially, the AHP method, originally proposed by Saaty (1980), simply requires decision makers to make a number of pairwise comparison judgments, with the idea that making preference judgments between pairs of alternatives is a rather natural and easy process. Starting from these pairwise comparisons, AHP then provides an objective procedure to compute the final weights (the so-called priorities in the AHP language), provided that the judgments expressed by decision makers give rise to consistent preferences.

More precisely, let us assume that we can get a group of museums' decision makers (for example museums' managers or other experts) to express their judgements on the relative importance of the $N = 4$ BSC perspectives:

1. Customer perspective
2. Internal process perspective
3. Innovation and learning perspective
4. Financial perspective

by asking them to focus on all the pairs of BSC perspectives:

1-2: Customer vs Internal process
1-3: Customer vs Innovation and learning
1-4: Customer vs Financial
2-3: Internal process vs Innovation and learning
2-4: Internal process vs Financial
3-4: Innovation and learning vs Financial

and indicate for each pairwise comparison: (i) which is the perspective they consider more important for the museums performance and (ii) which is the intensity of their preference, measured with the following fundamental scale adopted in the AHP approach (see, for example, Saaty 1990):

1 = Equal importance
3 = Moderate importance
5 = Essential or strong importance

7 = Very strong importance
9 = Extreme importance.

The results of the comparisons can be synthesized in the so-called pairwise comparison matrix:

$$A = \begin{pmatrix} 1 & a_{12} & a_{13} & a_{14} \\ a_{21} & 1 & a_{23} & a_{24} \\ a_{31} & a_{32} & 1 & a_{34} \\ a_{41} & a_{42} & a_{43} & 1 \end{pmatrix} \tag{6}$$

where by construction $a_{ji} = 1/a_{ij} \;\; \forall\, i, j$ (reciprocal property axiom) and A is a square matrix with dimension equal to the number N of alternatives compared (in our case, $N = 4$). Hence, for instance, if the decision maker prefers the Customer perspective to the Internal process with an intensity of 3, we associate the value of 3 to the comparison *Customer–Internal process* and its reciprocal (1/3) to the comparison *Internal process–Customer*.

However, AHP experiences have shown that a crucial issue regarding the pairwise comparison matrix is that often the decision makers' answers are not consistent, in the sense that they do not always satisfy the following transitivity property:

$$a_{ij} \cdot a_{jk} = a_{ik} \qquad \forall i, j, k. \tag{7}$$

Clearly, consistency is related to the coherence of the answers in the managers' pairwise comparisons. For example, if the customer perspective is evaluated as more important than the internal process perspective with intensity 3 and the internal process perspective is judged more important than the innovation and learning perspective with intensity 3, then consistency requires that the customer perspective is extremely more important than the innovation and learning perspective, with intensity 9.

In order to check the consistency of the pairwise comparison matrix A associated with a museum's expert, we can compute the consistency index (CI), defined as follows (see, e.g., Saaty 1990):

$$CI = \frac{\lambda_{\max} - N}{N - 1}, \tag{8}$$

where $\lambda_{\max}$ is the maximum eigenvalue of A. A perfectly consistent matrix should have a null consistency index, even if some inconsistency may be tolerated.

Usually, the value of the consistency index is compared to an "average" value, computed through a large set of randomly generated matrices and dependent on the matrix size N (see, e.g., Saaty 1990), with the so called consistency ratio (CR):

$$\mathrm{CR} = \frac{\mathrm{CI}}{\mathrm{RI}}, \tag{9}$$

where for the case $N = 4$ the random index RI is equal to 0.9. A value $\mathrm{CR} < 0.1$ is generally tolerated and associated with a "sufficiently" consistent matrix, while a higher value signals the presence of inconsistent judgements.

Generally, when the judgements are not consistent one should get back to the decision maker and ask him/her to review them. However, it is not always possible to ask

managers for another interview, and this is especially true in the museum sector. In these cases, it is possible to resort to one of the methods suggested in the literature to reduce the inconsistency of judgements or even to achieve consistency by properly modifying some elements of the pairwise comparison matrix; for a review of these methods see for example Pereira and Costa (2015). For instance, a convenient algorithm that enables to achieve consistency for an inconsistent pairwise comparison matrix is proposed in Benítez et al. (2011), which uses a linearization technique to provide the closest (in the Frobenius matrix norm) consistent matrix using an orthogonal projection in a linear space.

From the pairwise comparison matrix A, it is possible to determine the relative weights assigned to the various alternatives, called priorities in the AHP terminology; the priority vector coincides with the principal eigenvector of A, i.e., the eigenvector associated with the maximum eigenvalue of A, normalized so that its components sum to unity (see, e.g., Saaty 2003). In our case, these weights could be directly used to compute the weighted average of the four BSC scores obtained at the first stage, and we could take this as the overall performance score.

However, the management of museum institutions, especially public and nonprofit museums, often have multiple goals of different nature, and consequently museums' managers may have quite different interests regarding the museums activities and may attribute quite different weights to the various BSC perspectives. For example, curators will likely emphasize the importance of the customer and internal process perspectives, while financial managers will probably stress the relevance of the financial perspective. Therefore it will be necessary to interview managers and other experts in different positions. On the other hand, when multiple decision makers having different points of view are interviewed they may well express conflicting preferences.

Actually, the literature on AHP group decision making usually suggests that decision makers should review their preference judgments, even working together, hoping to reach a consensus (see for example Saaty and Vargas 2007, p. 923). However, in practice, this remedy is not always practicable. Furthermore, it could be reductive to force different museum's decision-makers to condense their opinions into a single collective judgment, when the differences come from different value systems.

On the other hand, the literature also suggests some ways to aggregate the individual preferences, either aggregating the judgments and obtaining a unique pairwise comparison matrix for the group, or aggregating the priorities; for a review on various group aggregation methods in the AHP framework, see (Ossadnik et al. 2016). In this regard, the most widely used aggregation tool is the geometric mean of the individual judgments; this is proved to be the best aggregation procedure for AHP since it allows to preserve the reciprocal property of the pairwise comparison matrix (see Aczél and Saaty 1983).

Nevertheless, when there are strong disagreement among the different decision-makers, the individual judgments might simply be too dispersed to be meaningfully aggregated in a unique group preference. In such a situation, the geometric mean is no longer representative of a group preference, but the literature does not provide many

other remedies, the most common approach being the suggestion to make decision makers review their preferences.

In order to measure the dispersion of individual judgments and test if they can meaningfully be aggregated using the geometric mean, Saaty and Vargas (2007) developed a specific statistical test called geometric dispersion test. More precisely, this is "*a statistical test to decide if the dispersion of a group can be considered as larger than usual, i.e., that the probability of obtaining the value of the sample geometric dispersion of the group is greater than a pre-specified significance level (e.g., 5%) in the distribution of the group geometric dispersion*" (Saaty and Vargas 2007, p. 922). When this test rejects the null hypothesis, there is a wide dispersion of managers' judgments that precludes the aggregation of their preferences with the geometric mean.

5 Applying AHP to MUVE museums

In order to verify the applicability of the AHP methodology to determine the weights to aggregate the BSC perspective scores, we interviewed five managers who are currently working in the municipal museums of Venice as field experts. The managers interviewed come from different functional areas: Administration (let these managers be denoted by D1 and D2), Directorate (D3), Communication and Business Development (D4), Security (D5).

With the aim of eliciting the preferences of these museums' managers, we first illustrated the meaning of the four BSC perspectives to each manager and then, as suggested by AHP, we asked her/him to compare the perspectives in pairs: [1-2] customer vs internal process, [1-3] customer vs innovation and learning, [1-4] customer vs financial, [2-3] internal process vs innovation and learning, [2-4] internal process vs financial, [3-4] innovation and learning vs financial, based on the relative importance she/he attached to the different dimensions of the museums performance.

Figure 2 displays the judgments (importance and intensity) given by each of the museums' experts interviewed. For example, with reference to the comparison *customer–internal process*, managers D1, D2 and D5 evaluate the customer perspective more important than the internal process perspective with an intensity equal to 3, while D4 gives the preference to the Internal process perspective with an intensity equal to 5 and D3, on the other hand, considers the two perspectives equally important.

We have also built the pairwise comparison matrices associated with the museum's managers preferences; they are reported in Table 9. The matrices were obtained from the judgments on the pairs of perspectives, so as to satisfy the reciprocal property $a_{ji} = 1/a_{ij}\ \forall\, i, j$.

In order to assess the consistency of managers' judgments, we have computed the consistency index CI and the consistency ratio CR; the results for all managers are reported in Table 10, together with the maximum eigenvalue $\lambda_{\max}$ of the pairwise comparison matrices. As often occurs in decision-making problems where several pairwise comparisons are made, we find some inconsistency in the judgments expressed. Actually, the CR value is higher than the tolerance value 0.1 for all decision makers D1–D5, and this value is especially high for D5.

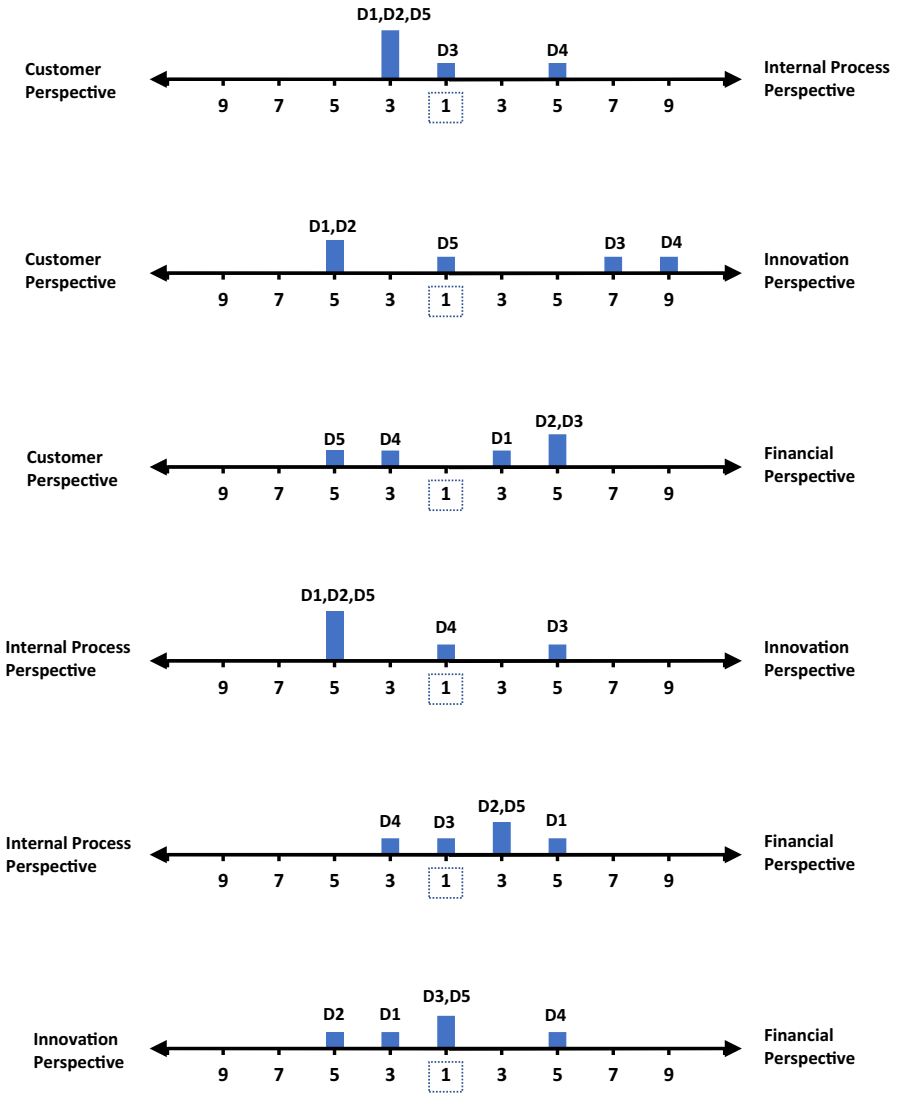

Fig. 2 Preferences (importance and intensity) expressed by museums' decision makers D1–D5 in the pairwise comparisons

Of course, the inconsistency of the preference judgements prevents us from using the final weights (the AHP priorities) that can be obtained with the AHP methodology just as they are. Anyway, for completeness, we have computed the priorities for all experts (see Table 11), as the eigenvector associated with the maximum eigenvalue of the pairwise comparison matrix, normalized so that its components sum to unity. As can be expected, we may observe considerable differences in the weights assigned to the four BSC perspectives by the different managers. Moreover, if we take into account the functional area of the experts interviewed, we can quite faithfully trace back these

Table 9 Pairwise comparison matrices for museums' decision makers D1–D5

	Customer	Internal process	Innovation and learning	Financial
D1				
Customer	1	3	5	1/3
Internal process	1/3	1	5	1/5
Innovation and learning	1/5	1/5	1	1/3
Financial	3	5	3	1
D2				
Customer	1	3	5	1/5
Internal process	1/3	1	5	1/3
Innovation and learning	1/5	1/5	1	1/5
Financial	5	3	5	1
D3				
Customer	1	1	1/7	1/5
Internal process	1	1	1/5	1
Innovation and learning	7	5	1	1
Financial	5	1	1	1
D4				
Customer	1	1/5	1/9	3
Internal process	5	1	1	3
Innovation and learning	9	1	1	5
Financial	1/3	1/3	1/5	1
D5				
Customer	1	3	1	5
Internal process	1/3	1	5	1/3
Innovation and learning	1	1/5	1	1
Financial	1/5	3	1	1

Table 10 Maximum eigenvalue (λ_{max}), consistency index CI and consistency ratio CR for the pairwise comparison matrices of museums' decision makers D1–D5

	λ_{max}	CI	CR
D1	4.5747	0.1916	0.2152
D2	4.4956	0.165	0.1856
D3	4.3541	0.1180	0.1326
D4	4.3957	0.1319	0.1482
D5	5.3880	0.4627	0.5198

differences to the functional area, which highlights the likely existence of different systems of values. For example, we may see that D1 and D2, from administration, give the highest weight to the financial perspective, while the experts from the other functional areas give a higher weight to other BSC perspectives.

On the other hand, it was not possible to make further interviews to the museums' managers, so that we decided to resort to the algorithm of Benítez et al. (2011) to

Table 11 AHP priorities of the BSC perspectives for museums' decision makers D1–D5. The last row shows the priorities obtained with the geometric means matrix G

Museums' manager	Customer	Internal process	Innovation and learning	Financial
D1	0.2739	0.1526	0.0707	0.5027
D2	0.2403	0.1530	0.0541	0.5525
D3	0.0808	0.1429	0.4723	0.3039
D4	0.1002	0.3516	0.4758	0.0724
D5	0.4203	0.2215	0.1527	0.2054
DG	0.2346	0.2388	0.2102	0.3164

obtain the closest (in the Frobenius matrix norm) consistent matrices and to compute consistent priorities from them.

In the following, we compare the consistent matrices $\hat{A}_{Dk}$ thus obtained to the original matrices A_{Dk} derived from the judgements actually expressed by the various managers Dk, with $k = 1, \ldots, 5$:

$$A_{D1} = \begin{pmatrix} 1.00 & 3.00 & 5.00 & 0.33 \\ 0.33 & 1.00 & 5.00 & 0.20 \\ 0.20 & 0.20 & 1.00 & 0.33 \\ 3.00 & 5.00 & 3.00 & 1.00 \end{pmatrix} \qquad \hat{A}_{D1} = \begin{pmatrix} 1.00 & 1.97 & 4.40 & 0.58 \\ 0.51 & 1.00 & 2.24 & 0.29 \\ 0.23 & 0.45 & 1.00 & 0.13 \\ 1.73 & 3.41 & 7.62 & 1.00 \end{pmatrix}$$

$$A_{D2} = \begin{pmatrix} 1.00 & 3.00 & 5.00 & 0.20 \\ 0.33 & 1.00 & 5.00 & 0.33 \\ 0.20 & 0.20 & 1.00 & 0.20 \\ 5.00 & 3.00 & 5.00 & 1.00 \end{pmatrix} \qquad \hat{A}_{D2} = \begin{pmatrix} 1.00 & 1.52 & 4.40 & 0.45 \\ 0.66 & 1.00 & 2.89 & 0.29 \\ 0.23 & 0.35 & 1.00 & 0.10 \\ 2.24 & 3.41 & 9.84 & 1.00 \end{pmatrix}$$

$$A_{D3} = \begin{pmatrix} 1.00 & 1.00 & 0.14 & 0.20 \\ 1.00 & 1.00 & 0.20 & 1.00 \\ 7.00 & 5.00 & 1.00 & 1.00 \\ 5.00 & 1.00 & 1.00 & 1.00 \end{pmatrix} \qquad \hat{A}_{D3} = \begin{pmatrix} 1.00 & 0.61 & 0.17 & 0.27 \\ 1.63 & 1.00 & 0.27 & 0.45 \\ 5.92 & 3.64 & 1.00 & 1.63 \\ 3.64 & 2.24 & 0.61 & 1.00 \end{pmatrix}$$

$$A_{D4} = \begin{pmatrix} 1.00 & 0.20 & 0.11 & 3.00 \\ 5.00 & 1.00 & 1.00 & 3.00 \\ 9.00 & 1.00 & 1.00 & 5.00 \\ 0.33 & 0.33 & 0.20 & 1.00 \end{pmatrix} \qquad \hat{A}_{D4} = \begin{pmatrix} 1.00 & 0.26 & 0.20 & 1.32 \\ 3.87 & 1.00 & 0.76 & 5.10 \\ 5.10 & 1.32 & 1.00 & 6.71 \\ 0.76 & 0.20 & 0.15 & 1.00 \end{pmatrix}$$

$$A_{D5} = \begin{pmatrix} 1.00 & 3.00 & 1.00 & 5.00 \\ 0.33 & 1.00 & 5.00 & 0.33 \\ 1.00 & 0.20 & 1.00 & 1.00 \\ 0.20 & 3.00 & 1.00 & 1.00 \end{pmatrix} \qquad \hat{A}_{D5} = \begin{pmatrix} 1.00 & 2.28 & 2.94 & 2.24 \\ 0.44 & 1.00 & 1.29 & 0.98 \\ 0.34 & 0.77 & 1.00 & 0.76 \\ 0.45 & 1.02 & 1.32 & 1.00 \end{pmatrix}$$

The matrices $\hat{A}_{Dk}$ are perfectly consistent, so that their maximum eigenvalue is equal to $N = 4$ for all of them, and consequently both their consistency index CI and their consistency ratio CR are equal to 0.

The final weights (AHP priorities) obtained with the consistent matrices $\hat{A}_{Dk}$ for all managers are exhibited in Table 12. It is interesting to note that the final weights

Table 12 AHP priorities of the BSC perspectives determined with the AHP consistent matrices $\hat{A}_{Dk}$, for museums' decision makers D1–D5. The last row shows the priorities obtained with the geometric means matrix $\hat{G}$

Museums' manager	Customer	Internal process	Innovation and learning	Financial
D1	0.2884	0.1465	0.0655	0.4995
D2	0.2428	0.1592	0.0552	0.5428
D3	0.0821	0.1335	0.4857	0.2986
D4	0.0932	0.3610	0.4750	0.0708
D5	0.4493	0.1971	0.1527	0.2009
DG	0.2363	0.2325	0.2081	0.3232

obtained with the consistent matrices are actually very close to the weights coming from the original pairwise matrices.

Once obtained consistent pairwise comparison matrices, we should aggregate them in order to get a unique matrix and a unique vector of final weights. However, from Fig. 2 it is evident that the judgments of the museum's decision makers interviewed are very dispersed. We may even observe that for none of the 6 comparisons the preference is given to the same perspective by all the experts interviewed. As a consequence, if we compute the priorities with the geometric means matrix (see the last row of Tables 11, 12), we are not far from an equal weighting system. Here an important role again is likely to be played by the managers' specific areas of expertise and their educational and cultural background.

The dispersion of the preference judgments expressed by D1–D5 is confirmed by the results of the Saaty–Vargas geometric dispersion test (Saaty and Vargas 2007) mentioned in the previous section. Indeed, we computed the geometric dispersion test for all the 6 comparison pairs, both for the original preference matrices A_{Dk} and for the closest consistent matrices $\hat{A}_{Dk}$ obtained with the linearization technique; the results are shown in Table 13, which reports the values of the geometric dispersion test and the associated p value for all comparisons. Using the usual 5% significance level, a p value higher than 0.05 means that a significant dispersion exists among the judgments of D1–D5; we can note that this occurs for all pairs of the BSC perspectives compared, both for the original and for the consistent matrices.

In the end, the wide dispersion exhibited by the experts' judgements does not advise to aggregate them in a unique vector of weights representative of a homogeneous group, even if we can easily compute the matrices of geometric means:

$$G = \begin{pmatrix} 1.00 & 1.40 & 0.83 & 0.72 \\ 0.71 & 1.00 & 1.90 & 0.58 \\ 1.20 & 0.53 & 1.00 & 0.80 \\ 1.38 & 1.72 & 1.25 & 1.00 \end{pmatrix} \qquad \hat{G} = \begin{pmatrix} 1.00 & 1.02 & 1.14 & 0.73 \\ 0.98 & 1.00 & 1.12 & 0.72 \\ 0.88 & 0.90 & 1.00 & 0.64 \\ 1.37 & 1.39 & 1.55 & 1.00 \end{pmatrix}$$

where G and $\hat{G}$ denote the geometric means matrices obtained from the original and from the closest consistent pairwise comparison matrices, respectively.

Table 13 Values of the geometric dispersion test and p values for the preference judgments of decision makers D1–D5; the values are reported for each pair of perspectives compared, both for the original preference matrices (columns 2–3) and the closest consistent matrices (columns 4–5)

Perspectives compared	Original geometric dispersion	p value	Consistent geometric dispersion	p value
Customer—Internal process	2.4933	0.2647	2.1156	0.1324
Customer—Innovation and learning	4.5239	0.8634	4.3249	0.8320
Customer— Financial	3.8218	0.7259	1.9784	0.0934
Internal process—Innovation and learning	3.1861	0.5304	2.0443	0.1114
Internal process—Financial	2.3934	0.2272	2.4780	0.2589
Innovation and learning—Financial	2.4780	0.2589	3.9536	0.7577

As we have recalled in the previous section, in this case the literature on group decision making within AHP suggests that decision makers should revise their preference judgments, if possible. Since, in our case, this was not possible, we had to devise a different expedient.

To this aim, in the next section we present a model that takes into account the preference judgements collected from the museums' experts in a "looser" way and combines the AHP methodology with a further DEA–BSC model.

6 A DEA–BSC–AHP three-system model for museum evaluation

In this section, we propose a DEA–BSC–AHP three-system model that takes into account the preference judgements expressed by all the museums' experts interviewed in order to define a proper set of weights restrictions in a DEA–BSC model. This allows us to consider the judgements of all the experts interviewed, no matter how dispersed they are.

More precisely, we will use the AHP preferences to determine weights restrictions to be applied to a second-stage DEA model that computes an overall performance indicator; the idea is still to synthesize the four BSC performance scores obtained with the first-stage DEA models presented in Sect. 3.1 using a different DEA–BSC model. Graphically, we can depict the flowchart of this three-system model as in Fig. 3.

Unlike the DEA model adopted in Basso et al. (2018) and recalled in Sect. 3.2, that applies the virtual weights restrictions methodology, for the weights restrictions of the DEA–BSC–AHP three-system model we adopt the so called assurance regions (AR) approach (Thompson et al. 1986). Hence, the weights restrictions derived by the judgements expressed by the museums' experts provide an alternative way to restrict the weights of the second-stage DEA–BSC model.

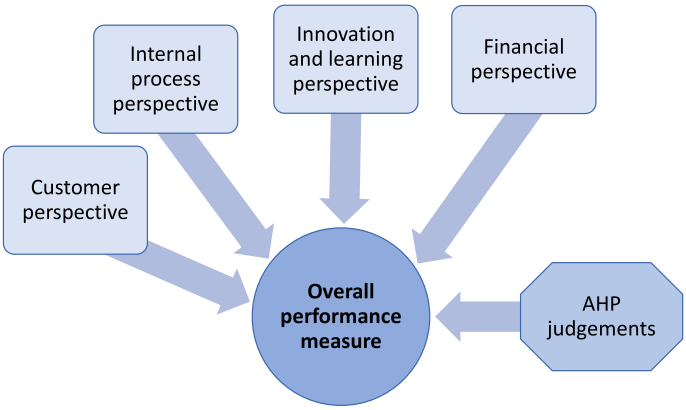

Fig. 3 Flow chart of the DEA–BSC–AHP three-system model

Indeed, this is the simplest and most natural way to define proper weights restrictions derived from the AHP pairwise comparison matrices of all experts, as the preferences are obtained by directly asking experts to give a judgement on the comparison between all pairs of BSC perspectives. It seems therefore natural to resort to restrictions imposed on the ratios between the weights of the output variables, just as suggested by the type I assurance region approach.

More specifically, type I assurance regions consider lower and upper bounds on the ratio between the weights of (either the inputs or) the outputs:

$$L^Y_{r'r} \leq \frac{u_r}{u_{r'}} \leq U^Y_{r'r} \qquad r, r' = 1, \ldots, t, \tag{10}$$

where $L^Y_{r'r}$ and $U^Y_{r'r}$ are the bounds of the output weights ratio $\frac{u_r}{u_{r'}}$. Notice that for $r' = r$ the lower bound may take any non negative value lower than or equal to 1 and the upper bound may take any value greater than or equal to 1 and the restrictions are redundant. Of course, a lower bound equal to 0 corresponds to the case without a limitation from below.

In our model, the restrictions on the output weights are associated with the four BSC perspectives performance scores obtained with the four DEA–BSC models of the first stage, while there are no restriction on the (unique) input weight.

Let us now see how we can derive proper lower and upper bounds for the AR restrictions (10) using the judgements of a group of M experts resulting from the interviews carried out. We have seen that from these judgements we can define a set of pairwise comparison matrices A_{Dk}, one for each expert Dk ($k = 1, \ldots, M$); let us assume that these matrices are (almost) consistent.

Let us focus on a pair of BSC perspective scores, say (r, r'); for example (Customer perspective, Internal process perspective). The idea is to define a range for the ratio between the weights of outputs r and r' such that it includes the results of the judgements expressed by all the M experts, no matter how different they are. Of course, the

more the judgements are similar, the narrower the range of values, while the more the judgements are dispersed, the wider the resulting range.

In particular, we propose to compute the bounds on the range for the ratio of the output weights as follows:

$$L^Y_{r'r} = (1+\alpha)^{-1} \min_{Dk} A_{Dk}(r, r') \qquad r, r' = 1, \ldots, 4 \tag{11}$$

$$U^Y_{r'r} = (1+\alpha) \max_{Dk} A_{Dk}(r, r') \qquad r, r' = 1, \ldots, 4 \tag{12}$$

where $\alpha \geq 0$ is a parameter that slightly widens the range with respect to the minimum and maximum intensity of the preferences expressed by the experts. Therefore, the constraints added to the second-stage DEA model can be written as follows:

$$(1+\alpha)^{-1} \min_{Dk} A_{Dk}(r, r') \leq \frac{u_r}{u_{r'}} \leq (1+\alpha) \max_{Dk} A_{Dk}(r, r') \qquad r, r' = 1, \ldots, 4. \tag{13}$$

Let us point out that the presence of the parameter α ensures that the interval $[L^Y_{r'r}, U^Y_{r'r}]$ have a width greater than 0 even when all the experts' judgements coincide.

A few papers in the DEA literature makes use of AHP to introduce AR restrictions on the weights of a DEA model, in different contexts. Takamura and Tone (2003) propose a similar approach for a site evaluation study concerning a relocation; in their paper, however, the AR bounds are derived from the AHP final weights of various evaluators instead than from the judgements directly expressed by the evaluators. Other contributions which use the AHP priorities of a set of decision makers to determine AR limitations are presented by Lee et al. (2012) for the photovoltaics industry, and by Kong and Fu (2012) to assess the performance of business schools.

7 Applying the DEA–BSC–AHP three-system model to MUVE museums

Let us now see how the DEA–BSC–AHP three-system model performs for the investigation on the MUVE museums.

The four BSC performance scores of stage 1 coincide with those obtained in Basso et al. (2018), reported in Table 6, while the results obtained with the DEA–BSC–AHP three-system model in stage 2 depend on the value chosen for the parameter α that affects the range of the weights restrictions.

Table 14 displays the performance scores obtained with the three-system model for values of α in the interval [0, 0.5]. The value $\alpha = 0$ corresponds to AR bounds obtained directly from the pairwise comparison matrices associate to the museums' managers D1–D5, without any enlargement; on the other hand, as the value of $\alpha = 0$ increases the range of the weights restrictions gets wider and wider. The results are compared with those obtained without AR restrictions on the output weights (see the last column of Table 14).

It is interesting to see how the introduction of the weights restrictions is effective in increasing the discriminatory power of the model. Indeed, from Table 14, we may

Table 14 Museums performance scores obtained at the second stage with the three-system DEA–BSC–AHP model in the analysis of the MUVE museums, for different values of the parameter α

Museums / α	0	0.1	0.2	0.3	0.4	0.5	Without AR bounds
Doge's Palace	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Correr Museum	0.8613	0.8713	0.8800	0.8876	0.8942	0.9001	1.0000
Glass Museum	0.9115	0.9164	0.9206	0.9243	0.9276	0.9305	1.0000
Ca' Rezzonico	0.7677	0.7808	0.7924	0.8027	0.8120	0.8203	0.9356
Ca' Pesaro	0.7598	0.7756	0.7895	0.8019	0.8131	0.8231	0.9642
Natural History Museum	0.6688	0.6834	0.6963	0.7078	0.7181	0.7274	0.8600
Mocenigo Palace	0.5030	0.5160	0.5274	0.5377	0.5468	0.5551	0.6775
Carlo Goldoni's House	0.8104	0.8238	0.8353	0.8454	0.8542	0.8620	1.0000
Clock Tower	0.9854	0.9867	0.9877	0.9886	0.9894	0.9901	1.0000
Lace Museum	0.7945	0.8037	0.8117	0.8186	0.8248	0.8302	1.0000
Fortuny Palace	0.6224	0.6351	0.6463	0.6562	0.6652	0.6733	0.7860

Table 15 Museums ranking obtained at the second stage with the three-system DEA–BSC–AHP model in the analysis of the MUVE museums, for different values of the parameter α

Museums / α	0	0.1	0.2	0.3	0.4	0.5	Without AR bounds
Doge's Palace	1	1	1	1	1	1	1
Correr Museum	4	4	4	4	4	4	1
Glass Museum	3	3	3	3	3	3	1
Ca' Rezzonico	7	7	7	7	8	8	8
Ca' Pesaro	8	8	8	8	7	7	7
Natural History Museum	9	9	9	9	9	9	9
Mocenigo Palace	11	11	11	11	11	11	11
Carlo Goldoni's House	5	5	5	5	5	5	1
Clock Tower	2	2	2	2	2	2	1
Lace Museum	6	6	6	6	6	6	1
Fortuny Palace	10	10	10	10	10	10	10

Table 16 Descriptive statistics of the results obtained at the second stage with the three-system DEA–BSC–AHP model in the analysis of the MUVE museums, for different values of the parameter α

α	Min score	Mean score	Median of score	Std.dev. score	N.efficient museums	Ranking variation
0.0	0.5030	0.7895	0.7945	0.1514	1	0.7727
0.1	0.5160	0.7993	0.8037	0.1471	1	0.7727
0.2	0.5274	0.8079	0.8117	0.1434	1	0.7727
0.3	0.5377	0.8155	0.8186	0.1402	1	0.7727
0.4	0.5468	0.8223	0.8248	0.1373	1	0.6818
0.5	0.5551	0.8284	0.8302	0.1347	1	0.6818
Without AR bounds	0.6775	0.9294	1.0000	0.1095	6	0.0000

observe that only the Doge's Palace is efficient when the weights restrictions are imposed, while the limiting case without AR restrictions (that can be obtained for $\alpha \to \infty$) sees as many as 6 efficient museums over 11.

In order to analyze the influence of α on the performance results of the museums, let us also present the ranking determined by the performance scores of Table 14, ranking displayed in Table 15. Interestingly, the ranking shows only a change between $\alpha = 0.3$ and $\alpha = 0.4$, and precisely a swap between the positions 7 and 8 (Ca' Rezzonico and Ca' Pesaro museums). Actually, while the performance scores vary with α, the other relative positions are not affected, even in the limiting case without AR restrictions. A few more changes in the ranking can be observed with respect to the results of the two-stage DEA–BSC model with virtual weights instead of AR restrictions (see Table 8).

For completeness, Table 16 summarizes some descriptive statistics on the performance scores: minimum, mean and median scores, standard deviation of the score

and number of efficient museums. In addition, the last column of the table reports the mean absolute variation of the ranking positions divided by 2, for the different values of α; the variation is computed with respect to the case without AR restrictions on the output weights.

8 Conclusion

In this paper, we propose a new two-stage model that can be used to evaluate and compare the performance of a set of museums. The first stage is modeled on the joint use of the BSC and DEA approaches as proposed in Basso et al. (2018) and enables the museums top management to measure the relative efficiency of museums along each perspective, thus allowing one to identify museum's strengths and weaknesses.

The second stage has the objective to synthesize the different performance results in a single recapitulatory indicator and this can be done in different ways. The idea pursued in the present paper is to try to make use of an AHP approach in order to combine the performance indicators obtained for the different BSC perspectives in the first stage according to the preferences of a group of museums experts.

The first attempt is to directly use an AHP approach, i.e., to weigh the DEA–BSC scores of the first stage with the AHP priorities computed from the judgements expressed by the experts. However, very often in the museums sector, experts with different backgrounds give completely different judgements, so deeply different that it is not possible to aggregate their preferences into a single one that is representative of the "group preference".

A different approach combines the AHP approach with a DEA model and tries to extract useful information on the weights from the set of preferences of all the experts interviewed, even when they are very dispersed. This information is used to impose proper restrictions (more precisely, AR restrictions) on the weights of a DEA model in which the outputs are give by the DEA–BSC scores of the first stage. This gives rise to a two-stage, three-system model that integrates DEA, BSC and AHP and provides an overall performance measure that takes all the perspectives into consideration simultaneously.

The overall performance indicator signals which of the museums analysed are managed in an optimal way and which of them could instead try to improve their management.

Funding Open access funding provided by Università Ca' Foscari within the CRUI-CARE Agreement.

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

Aczél, J., Saaty, T.L.: Procedures for synthesizing ratio judgments. J. Math. Psychol. **27**, 93–102 (1983)

Basso, A., Funari, S.: Measuring the performance of museums: classical and FDH DEA models. Rend. Stud. Econ. Quant. **2003**, 1–16 (2003)

Basso, A., Funari, S.: A quantitative approach to evaluate the relative efficiency of museums. J. Cult. Econ. **28**, 195–216 (2004)

Basso, A., Casarin, F., Funari, S.: How well is the museum performing? A joint use of DEA and BSC to measure the performance of museums. Omega. Int. J. Manag. Sci. **81**, 67–84 (2018)

Benítez, J., Delgado-Galván, X., Izquierdo, J., Pérez-García, R.: Achieving matrix consistency in AHP through linearization. Appl. Math. Model. **35**, 4449–4457 (2011)

Carvalho, P., Silva Costa, J., Carvalho, A.: The economic performance of Portuguese museums. Urban Public Econ. Rev. **20**, 12–37 (2014)

Cherchye, L., Moesen, W., Rogge, N., Van Puyenbroeck, T.: An introduction to 'benefit of the doubt' composite indicators. Soc. Ind. Res. **82**(1), 111–145 (2007)

Cooper, W.W., Seiford, L.M., Tone, K.: Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software. Kluwer Academic Publishers, Boston (2000)

Del Barrio, M.J., Herrero, L.C., Sanz, J.A.: Measuring the efficiency of heritage institutions: a case study of a regional system of museums in Spain. J. Cult. Herit. **10**, 258–268 (2009)

Del Barrio, M.J., Herrero, L.C.: Evaluating the efficiency of museums using multiple outputs: Evidence from a regional system of museums in Spain. Int. J. Cult. Policy **20**, 221–238 (2014)

Del Barrio-Tellado, M.J., Herrero-Prieto, L.C.: Modelling museum efficiency in producing inter-reliant outputs. J. Cult. Econ. **43**, 485–512 (2019)

Haldma, T., Lääts, K.: The balanced scorecard as a performance management tool for museums. In: Gregoriou, G.N., Finch, N. (eds.) Best Practices in Management Accounting, pp. 232–52. Palgrave Macmillan, London (2011)

Kaplan, R.S., Norton, D.P.: Transforming the Balanced Scorecard from performance measurement to strategic management Part I. Account. Horiz. **15**, 87–104 (2001)

Kong, W.-H., Fu, T.-T.: Assessing the performance of business colleges in Taiwan using data envelopment analysis and student based value-added performance indicators. Omega. Int. J. Manag. Sci. **40**, 541–549 (2012)

Lee, A.H.I., Lin, C.Y., Kang, H.-Y., Lee, W.H.: An integrated performance evaluation model for the photovoltaics industry. Energies **5**, 1271–1291 (2012)

Lovell, C.A.K., Pastor, J.T.: Radial DEA models without inputs or without outputs. Eur. J. Oper. Res. **118**(1), 46–51 (1999)

Mairesse, F., Vanden, Eeckaut P.: Museum assessment and FDH technology: towards a global approach. J. Cult. Econ. **26**, 261–86 (2002)

Marcon, G.: La gestione del museo in un'ottica strategica: L'approccio della Balanced Scorecard. In: Sibilio Parri, B. (ed.) Misurare e comunicare i risultati L'accountability del museo, pp. 21–56. Franco Angeli, Milano (2004)

Ossadnik, W., Schinke, S., Kaspar, R.H.: Group aggregation techniques for analytic hierarchy process and analytic network process: a comparative analysis. Group Decis. Negot. **25**, 421–457 (2016)

Pereira, V., Costa, H.-G.: Nonlinear programming applied to the reduction of inconsistency in the AHP method. Ann. Oper. Res. **229**, 635–655 (2015)

Pignataro, G.: Measuring the efficiency of museums: a case study in Sicily. In: Rizzo, I., Towse, R. (eds.) The Economics of Heritage: A Study in the Political Economy of Culture in Sicily, pp. 65–78. Edward Elgar Publishing Ltd, New York (2002)

Saaty, T.L.: The Analytic Hierarchy Process. McGraw-Hill, New York (1980)

Saaty, T.L.: How to make a decision: the analytic hierarchy Process. Eur. J. Oper. Res. **48**, 9–26 (1990)

Saaty, T.L.: Decision-making with the AHP: why is the principal eigenvector necessary. Eur. J. Oper. Res. **145**, 85–91 (2003)

Saaty, T.L., Vargas, L.G.: Dispersion of group judgments. Math. Comput. Model. **46**, 918–925 (2007)

Taheri, H., Ansari, S.: Measuring the relative efficiency of cultural-historical museums in Tehran: DEA approach. J. Cult. Herit. **14**, 431–438 (2013)

Takamura, Y., Tone, K.: A comparative site evaluation study for relocating Japanese government agencies out of Tokyo. Soc. Econ. Plan. Sci. **37**, 85–102 (2003)

Thompson, R.G., Singleton Jr., F.D., Thrall, R.M., Smith, B.A.: Comparative site evaluations for locating a high-energy physics lab in Texas. Interfaces **16**, 35–49 (1986)

Wei, T.L., Davey, H., Coy, D.: A disclosure index to measure the quality of annual reporting by museums in New Zealand and the UK. J. Appl. Account. Res. **9**, 29–51 (2008)

Wong, Y.H.B., Beasley, J.E.: Restricting weight flexibility in data envelopment analysis. J. Oper. Res. Soc. **41**, 829–835 (1990)

Zorloni, A.: Designing a strategic framework to assess museum activities. Int. J. Arts Manag. **14**, 31–47 (2012)

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