European Journal of Operational Research 297 (2022) 795–816

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Invited Review

# Integrating multiple criteria decision analysis and production theory for performance evaluation: Framework and review ☆

Harald Dyckhoff[a,*], Rainer Souren[b]

[a] School of Business and Economics (former Chair of Business Theory, Sustainable Production and Industrial Control), RWTH Aachen University, Templergraben 64, 52056 Aachen, Germany

[b] Group of Sustainable Production and Logistics Management, Ilmenau University of Technology, Postbox 100565, 98684 Ilmenau, Germany

A R T I C L E I N F O

*Article history:*
Received 9 December 2020
Accepted 26 May 2021
Available online 6 June 2021

*Keywords:*
Multiple criteria analysis
Production theory
Value function
Data envelopment analysis (DEA)
Life cycle assessment (LCA)

A B S T R A C T

Accounting, life cycle assessment (LCA) and data envelopment analysis (DEA) are examples of various research areas that independently develop and apply diverse methodologies to evaluate performance. Though, many methods have in common that the results to be assessed are mainly determined by the inputs and outputs of the activities which are to be evaluated. Based on both production and decision theory, our comprehensive framework integrates and systematically distinguishes specific types of production-based performance assessment. It allows to examine and categorise the existing literature on such approaches. Our review focuses on sources which explicitly apply concepts or methods of multiple criteria decision analysis (MCDA). We did not find any elaborated methodology that fully integrates MCDA with production theory. At least, a basic approach to multicriteria performance analysis, which generalises the methodology of data envelopment analysis, appears to be well-grounded on production theory. It was already presented in this journal in 2001 and has rarely been noticed in the literature until now. A short overview outlines its recent insights and main findings. A key finding is that a category mistake prevails among well-known methodologies of efficiency measurement like DEA. It may imply invalid empirical results because the inputs and outputs of production processes are confused with resulting impacts destroying or creating values (to be minimised or maximised, respectively). We conclude by defining open problems and by indicating prospective research directions.

© 2021 The Author(s). Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license
(http://creativecommons.org/licenses/by-nc-nd/4.0/)

## 1. Diverse production-based methodologies of performance evaluation

Performance evaluation methods are used to assess relevant results of human activities. Such results are also called consequences, effects, impacts, or outcomes. Methods of business administration and economics are mostly concerned with financial results on markets, or, in case of traditional cost-benefit-analysis, they measure how much individuals would pay for the results (willingness-to-pay). Since non-financial results are of growing importance, further methodologies have been developed to evaluate environmental and social aspects of performance, too. Their approaches may differ to such an extent that they cannot be easily compared with each other or with monetary-based methods.

*Life cycle assessment* (LCA) and *data envelopment analysis* (DEA) are amongst these diverse methodologies. However, LCA and DEA have one thing in common with *management accounting*: The results to be assessed are mainly determined by the observed inputs and outputs of the activities which are to be evaluated. Therefore, these performance evaluation methodologies can be called *production-based*. In the case of cost and revenue accounting, the inputs and outputs of a production process are usually valued at their market prices when purchased or sold.

In LCA, the assessment of ecological impacts is built on an inventory of all relevant inputs and outputs concerning the product under consideration during its manufacture, use and subsequent disposal (Guinee, et al., 2011). For example, *Figure 1* shows six main impact categories from the BASF's *eco-efficiency analysis*, a well-established method in LCA practice (cf. the overview by Grosse-Sommer et al. (2020)). Three of them are input-related and three output-related. Within the last ($3^{rd}$) step of BASF's analysis, all six are aggregated into a single category called “total ecological impact”. The figure illustrates that, at first, different kinds of

☆ This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

* Corresponding author.
*E-mail addresses:* Dyckhoff@lut.rwth-aachen.de (H. Dyckhoff), Rainer.Souren@tu-ilmenau.de (R. Souren).

https://doi.org/10.1016/j.ejor.2021.05.046
0377-2217/© 2021 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

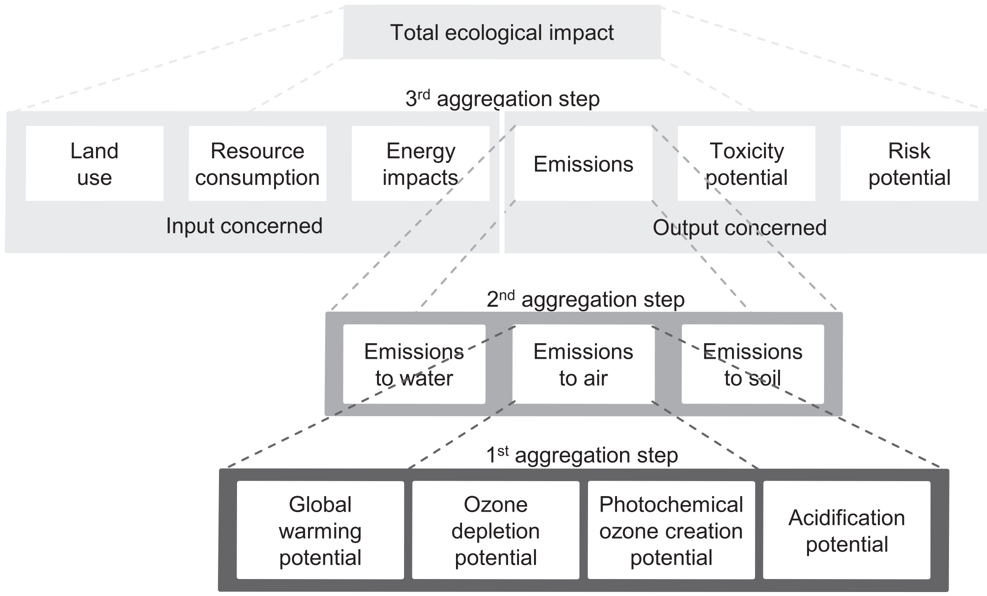


**Fig. 1.** Aggregation scheme of ecological performance criteria (cf. Dyckhoff, Quandel & Waletzke (2015), p. 1560)

emissions – usually classified as undesirable outputs – are aggregated hierarchically in two preceding steps. Eco-efficiency analysis is a particular instance of *multicriteria evaluation methods* which are based on input and output data of the considered processes.

This holds also true for DEA (cf. Doyle & Green (1993), Joro, Korhonen & Wallenius (1998), p. 963). A crucial question of its application is the selection and definition of the inputs and outputs which are relevant for the performance evaluation of the decision-making units (DMUs) at hand. Cook, Tone & Zhu (2014) state in a methodological review (p. 2):

> In summary, if the underlying DEA problem represents a form of ‘production process’, then ‘inputs’ and ‘outputs’ can often be more clearly identified. The resources used or required are usually the inputs and the outcomes are the outputs.
>
> If, however, the DEA problem is a general benchmarking problem, then the inputs are usually the ‘less-the-better’ type of performance measures and the outputs (…) the ‘more the-better’ type (…). DEA then can be viewed as a multiple-criteria evaluation methodology where DMUs are alternatives, and the DEA inputs and outputs are two sets of performance criteria where one set (inputs) is to be minimized and the other (outputs) to be maximized.

Each of these two seemingly unconnected alternatives – *production versus multicriteria perspective* – stated by Cook, Tone & Zhu (2014) poses its own difficulties. This can be illustrated by the example of assessing the sustainability performance of coal-fired power plants. Here, from a production technology point of view, undesired emissions like $CO_2$ are undoubtedly outputs, however, they need to be minimised from an ecological perspective. In addition, the production factor labour clearly constitutes an input; since labour input induces costs for the shareholders of the plant, it is minimised in business economics. Nevertheless, it represents at the same time employment which, in turn, constitutes a factor highly preferred by society which is thus to be maximised with regard to its stakeholders (cf. Iribarren & Vazquez-Rowe (2013), p. 125).

This constellation illustrates a fundamental conflict between the production and the multicriteria perspective. If both are considered as two mutually exclusive alternatives (as cited above), associated performance analyses will inevitably lead to completely different results, due to the possibility that an input or output must be optimised with opposite directions in view. This contradiction may call into question the empirical validity of the findings thus obtained.[1] Moreover, in case of a pure multicriteria perspective, a production theoretical foundation is lacking which would allow to envelop the measured input and output data of the considered DMUs by using assumptions like convexity or certain economies of scale in order to establish a set of additional production activities which are fictional, but nevertheless (technologically) possible (Wojcik, Dyckhoff, & Clermont, 2019).

To resolve this unsatisfactory contradiction found in the literature of performance evaluation, our review investigates two main questions: Is there a bridge that connects the production and the multicriteria perspective or even integrates both perspectives into a single *multicriteria production perspective* as a comprehensive framework of performance evaluation in such a way that the afore apparently antagonistic perspectives come to form ‘two sides of the same coin’? What would a theoretically sound approach which systematically integrates both perspectives into an encompassing methodical conception of *production-based multicriteria performance analysis* look like?

In response to these questions, the next section proposes a multicriteria production framework for performance evaluation. It integrates fundamental concepts from different disciplines, notably purposive rationality from sociology and multicriteria evaluation from decision theory with the transformation of input into output objects from production theory. Section 3 demonstrates that accounting, LCA, and DEA form specific types of methodologies classified by this framework. On this basis, as the first general review of this kind, Section 4 examines, categorises, and discusses the literature of Operations Research and Management Science about concepts and approaches combining multicriteria performance evaluation with production theory or with the production-based topics of accounting, LCA, or DEA. It emerges that relatively few sources try to bridge the gap between these different strands of research if compared to the extensive literature dealing with each strand alone. Furthermore, there seems to be only a sin-

[1] In an often-cited review of contributions to supplier evaluation and selection, Ho, Xu & Dey (2010), p. 22, asserted that this ambiguity is a limitation or drawback of DEA and conclude that “practitioners may be confused with input and output criteria”.

gle approach to multicriteria performance analysis which is well-grounded on production theory. Although it was already presented in this journal two decades ago (Dyckhoff & Allen, 2001), it has been noticed in the literature solely regarding the aspect of ecological efficiency measurement whereas the general multicriteria concept has been widely ignored until now. Section 5 outlines and further clarifies insights and key findings of corresponding recent literature with a focus on the data envelopment methodology of production-based multicriteria performance analysis. Most notably, it explains why the application of well-known methodologies like DEA often leads to the category mistake of confusing the inputs and outputs of production processes with resulting impacts which destroy or create values (to be minimised or maximised, respectively). Section 6 concludes by defining open problems and indicating prospective research directions that emerge from our review.

## 2. Comprehensive framework of performance evaluation

From a perspective of management accounting and control, performance evaluation is an instrument predominantly used for two ultimate purposes with typically conflicting conclusions (Demski & Feltham, 1976), namely either to facilitate and support decisions of a single decision maker or a team or else to support a principal in the process of influencing and controlling decisions of opportunistic agents in situations characterised by asymmetrically distributed information. In this section, we present different perspectives on performance evaluation, after having defined our understanding of it. By further developing ideas of Dyckhoff (2018) and Dyckhoff & Souren (2020), these perspectives will then be integrated into a comprehensive, theoretically founded framework, proposing generic guidelines for production-based performance assessment. It includes accounting, LCA, and DEA as particular instances of such methodologies, to be classified with respect to this framework in Section 3.

### 2.1. Different perspectives of performance evaluation

Performance analyses regularly assume that the acting individuals involved behave rationally. This fundamental assumption is first dealt with in more detail before we turn to the decision-theoretical and production-theoretical perspectives on performance evaluation.

#### 2.1.1. Purposive rationality as basic assumption

One century ago, the well-known sociologist and economist Max Weber (1921) wrote that “action is rationally oriented to a system of discrete individual ends when the ends, the means, and the secondary results are all rationally taken into account and weighted” (p. 21; cf. the translation in Weber, Henderson & Parsons (1964)). Weber (1921) distinguishes this *purposive rationality* from other categories of rationality, e.g. legality, legitimacy or morality, characterising a kind of *value rationality*.

Ends, means, and secondary results are three categories of criteria determining the rationality of an action or of an actor. The term ‘ends’ is used to name the purposes which constitute the original motives for the action considered in the situation at hand. The extent to which these main ends are achieved determines the *effectiveness* (*efficacy*, *effectivity*) of an action, whereas the consideration of the ends in relation to the means as well as to the secondary results appraises its *efficiency*. Effectiveness and efficiency form the two main categories regarding the quality of the decision being put into action. Such an analysis can be called *performance evaluation* (*assessment*) according to the pertinent literature on performance measurement which defines the latter as “the process of quantifying the efficiency and effectiveness of action” (Neely, Gregory & Platts (1995), p. 80). Performance evaluation comprises data from (exact) measurement, and in addition considers qualitative aspects if they are relevant.

The different types of ends E, means M, desirable secondary results D and undesirable ones U represent the relevant multiple criteria for the intended performance evaluation. Waste and emissions are examples of undesirable secondary outcomes, whereas a surprising discovery or invention made during an exploration process may be desirable. In general, the achievement of an end is desired and valued as a benefit whereas employing a means is undesired and is assessed as a cost factor.

Let $\mathcal{E}$ denote the *evaluation set*, i.e. those actions the performances of which are to be evaluated. If $\boldsymbol{b} = (\boldsymbol{b}^{\mathrm{E}};\boldsymbol{b}^{\mathrm{D}})$ describes the multi-dimensional values of all desirable results and $\boldsymbol{c} = (\boldsymbol{c}^{\mathrm{M}};\boldsymbol{c}^{\mathrm{U}})$ those of all undesirable ones, we call $\boldsymbol{b} = \boldsymbol{b}[a]$ the *benefits* and $\boldsymbol{c} = \boldsymbol{c}[a]$ the *costs* of action $a \in \mathcal{E}$. These different costs and benefits form the *multiple values destroyed or created* by the action in question: $\boldsymbol{v}[a] = (\boldsymbol{c};\boldsymbol{b})$. By definition, less costs and more benefits are preferred.

Thus, performance evaluation of actions can be regarded as a kind of non-monetary *cost/benefit-analysis* that generalises the monetary approaches of economics. Typically, some or all benefits and costs cannot be measured in financial terms or are even not measurable at all. If measurable, the scales of the different types of costs and benefits are regularly incommensurable *a priori*, particularly in cases of sustainability evaluations based on social and ecological criteria (cf. Figure 1). Therefore, non-financial performance evaluation constitutes a kind of multicriteria analysis in which the criteria cannot easily be aggregated into a one-dimensional measure of overall performance determining the *success* of an action (in contrast to e.g. the profit as difference of revenues and costs in financial accounting). Thus, eliciting success by employing methods developed by *multicriteria decision analysis* (MCDA), e.g. as an overall utility or value by multi-attribute utility/value theory (MAU/VT), may be helpful.

#### 2.1.2. Decision-theoretic perspective

Since the last quarter of the 20th century, the literature on MCDA has proposed and applied an abundance of approaches and methods for tackling decision-making problems with multiple objectives (cf. Wallenius et al. (2008), Cinelli et al. (2020)). To make them systematically usable for performance analysis, our structural framework is furthermore based on a decision-theoretic perspective. Its fundamental assumption says that a complex decision problem can be better analysed by decomposing it into five basic components:

(1) the set of alternatives (or option set) $\mathcal{A}$ as compilation of all actions $a$ that can be executed by the considered decision-maker (or DMU)
(2) the situation set $\mathcal{S}$ as compilation of all possible, uncertain scenarios $s$
(3) the result function $r(a;s)$ describing the consequences of action $a \in \mathcal{A}$ in situation $s \in \mathcal{S}$
(4) the relevant objectives regarding the consequences
(5) the preferences regarding the multiple objectives.

These components describe the basic decision model. As performance evaluation mostly analyses actions of the past, we focus on the special case of decisions under certainty, i.e. where the situation set consists of one element only. Hence $\mathcal{S}$ will be ignored in the following considerations.

Usually, the alternatives $a \in \mathcal{A}$ are only briefly denominated for clear identification. Instead, their results $r$ have to include all the information necessary for the valuation regarding the different objectives. They can generally be described in any way, even purely verbally. In contrast to this, and without lowering the level of generality very much, it can be assumed for each single objective that the extent to which it is fulfilled can be measured by means of an appropriate numerical scale (Roberts, 1979). Thus, for each individual objective, an associated one-dimensional value function[2] exists, usually unknown in advance. The objectives and value functions may not only be determined by one and the same decision maker alone but by different people, units, or stakeholders in situations known from game theory. All these individual functions together constitute a *multi-dimensional value function* $\boldsymbol{v} = \boldsymbol{w}(\boldsymbol{r}) = \boldsymbol{w}(\boldsymbol{r}(a))$ that represents the (generally conflicting) preferences for the different objectives. This value function has to be distinguished from an overall (utility or) value function $u = u(\boldsymbol{v})$ which would aggregate all the different multiple values on a one-dimensional numerical scale.

Methods of MCDA, like those of MAVT or multiple objective linear programming (MOLP), can help to gain and utilise information on the preferences regarding the values of the individual objectives as well as on the trade-off between them. According to our understanding, MCDA encompasses not only the analysis of the rationality of intended decision making for *actions in the future* (MCDM), but moreover also the assessment of the rationality (or 'performance') of already realised decisions – as well as the assessment of the people and units (DMUs) that have made these decisions –, usually judged by an (external) evaluator regarding the *results of past actions*.

#### 2.1.3. Production-theoretic perspective

Modern economic theory and DEA as a methodology for performance measurement strongly rely on the production theories of Koopmans (1951) and Shephard (1970). However, both use 'production', 'input' and 'output' as undefined basic terms. In the present paper, *production* is understood as value creation, i.e. as a process directed and controlled by human beings which transforms (input into output) objects with the intention of generating advantages that outweigh the disadvantages of the transformation. Positive values shall be created, and negative ones be destroyed to such an extent that all positive values consumed by the process as well as all negative values provoked by newly induced undesired results are more than offset.

Since the definition and selection of inputs and outputs is a critical issue of all production-based performance evaluation methodologies, a quote of Ragnar Frisch (1965) is helpful (p. 3):

> The term transformation indicates that there are certain things (goods or services) which enter into the process, and lose their identity in it, i.e. 'ceasing to exist' in their original form, while other things (goods or services) come into being in that they 'emerge' from the process. The first category may be referred to as 'production factors' (input elements), while the last-named category are referred to as 'products' (the output or resultant elements).

Hence, by definition, *input* enters into and *output* emerges out of the transformation process. Both the input of production factors and the output of products are understood as flows measured in time rates.

Apart from (desired) goods and services, the objects going into or emerging from the process may also turn out to be undesirable *bads*, which are assessed in an entirely different way than goods. A bad is an object whose possession is undesired so that one wants to get rid of it, which, however, requires some effort. Therefore, the output flow of bads is undesirable, e.g. in case of trim loss, whereas the input of bads is a desirable flow, e.g. of waste to be incinerated, because it decreases its stock (Dyckhoff & Allen, 2001).

[2] According to MAVT, the term 'value function' (and not 'utility function') is used here for decisions under certainty.

Production theory uses different notions for the production possibilities of a production system. Here, *technology* $\mathcal{T}$ comprises all activities $a$ that are *feasible in principle* because of the available knowledge as well as the deployable transformation processes. Apart from that, i.e. irrespective of the technology and in addition to it, the accessible activities are situationally narrowed and bounded by supply, sales and emission restrictions, all constituting a set $\mathcal{R}$. Therefore, the *really feasible* activities result from the intersection $\mathcal{P} = \mathcal{T} \cap \mathcal{R}$ of the activities that are both technologically feasible and satisfy the actual restrictions as well (Dyckhoff, 1992). The set $\mathcal{P}$ is called *production possibility set* (PPS).

Traditional production theories are limited to measurable inputs and outputs. Furthermore, they assume that each production activity $a \in \mathcal{P}$ can be completely described by process-specific *input quantities* $\boldsymbol{x}$ and *output quantities* $\boldsymbol{y}$. Such an activity $a$ can be identified by the representation $\boldsymbol{z} := (\boldsymbol{x}; \boldsymbol{y})$ of the related input and output quantities: $a \equiv \boldsymbol{z}$. Hence, there is no need to make a further distinction between the activity itself on the one hand, and the description of its inputs and outputs on the other. Thus, PPS $\mathcal{P}$ covers the *de facto* feasible production activities as vectors $\boldsymbol{z} = (\boldsymbol{x}; \boldsymbol{y})$ in a multi-dimensional space of real numbers, the dimensions of which are usually defined by the different types of inputs and outputs.

### 2.2. Integrating the different perspectives

Purposive rationality as well as decision and production theory form the building blocks and constitutive perspectives of production-based performance evaluation methods. Next, we compose them into the intended comprehensive framework. *Figure 2* outlines the basic structure of the latter as generic guideline for a recursive process with four strongly interdependent steps or components.

In the fundamental (first) step, the evaluation's subject and objectives as well as the decision field with the relevant alternatives $a \in \mathcal{E}$ which are to be assessed will be specified; this includes the goal and main purposes of a performance assessment as well as the determination of the relevant types of inputs and outputs combined with those types of resulting impacts which influence the selected objectives (*goal and scope determination*). In the second step, the input and output quantities $\boldsymbol{z}(a) = (\boldsymbol{x}; \boldsymbol{y})$ which influence the selected objectives have to be determined (*input/output inventory*). Based on that, all relevant outcomes $\boldsymbol{r} = \boldsymbol{r}(\boldsymbol{z})$ are assessed (*impact assessment*) and then valued by one or more performance measures $\boldsymbol{v} = \boldsymbol{w}(\boldsymbol{r})$ (*valuation*). The decision-maker or an external evaluator is responsible for the interpretation of the results in each step.

$\boldsymbol{r}(\boldsymbol{z})$ represents an *impact function*, whereas $\boldsymbol{w}(\boldsymbol{r})$ and $\boldsymbol{v}(\boldsymbol{z})$ are the respective *impact-related* and (transformation) *process-related value functions*. In general, these functions may be non-linear (cf. Sections 2.3 and 5). In fact, applied performance measurement methodologies mostly use linear functions, perhaps justified as an approximation of non-linear ones in the neighbourhood of the observed performance data.

Section 3 will show by way of example that specific methodologies – like accounting, LCA, and DEA – neither necessarily need to include all four steps of the generic structure in Figure 2 nor always consist of exactly two steps of (hierarchical) impact assessment and valuation: $\boldsymbol{v}(\boldsymbol{z}) = \boldsymbol{w}(\boldsymbol{r}(\boldsymbol{z}))$. Nonetheless, in cases of more steps or evaluation levels (cf. Figure 1), steps on the left (lower levels) are typically more often concerned with assessing objective or at least intersubjectively determinable facts, steps on the right (upper levels) more often deal with subjectively shaped values and preferences underlying the evaluation of these facts.

In any case, at the lowest levels of evaluation, the information that is relevant for each activity's performance consists of its specific inputs and outputs $\boldsymbol{z} = (\boldsymbol{x}; \boldsymbol{y})$ as well as of their corresponding

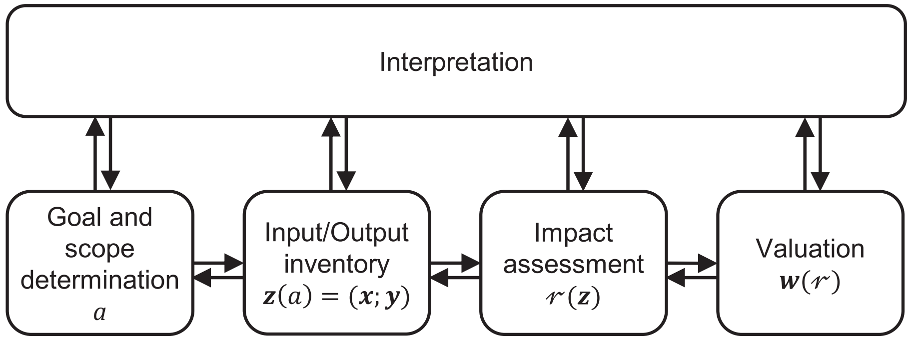


**Fig. 2.** Generic structure of performance evaluation (cf. Dyckhoff & Souren (2020), p. 8)

impacts $r(\boldsymbol{z})$. Regarding the impact assessment, it is important to note that one and the same input or output quantity can have different results simultaneously. For example, emissions of fluorocarbons into the air contribute to both global warming and to ozone depletion above Antarctica. At higher levels, the evaluation proceeds by using appropriate performance measures that, as a rule, condense the valuation further and further, thus allowing for a better comparison of the activities regarding their preferability. At the highest levels, the performance is represented by a few *key performance indicators* (KPI) only, or in extreme cases by a single, one-dimensional top KPI measuring the overall success.

The attributes representing the performance criteria – often classified into costs and benefits: $\boldsymbol{v} = (\boldsymbol{c}; \boldsymbol{b})$ – are usually defined by those ends E, means M as well as desirable and undesirable secondary results D and U which the evaluator chooses according to their relevance. Both, the results $r$ and the performance values $\boldsymbol{v}$ of the actions $a \in \mathcal{E}$ to be evaluated can generally be described in any way. By contrast, performance measurement requires a kind of quantification of all costs and benefits, each measurable on a well-defined numerical scale, usually at least an interval scale – maybe even a ratio scale –, but possibly an ordinal scale (French (1986), Ch. 9).

### 2.3. Illustrative example

The following numerical example is adapted from Dyckhoff (2018) and extended such that it is in line with the guidelines illustrated by Figure 2 and furthermore includes non-linear impact and value functions. The sustainability performance of several similar production units (DMUs) is to be evaluated regarding economic, social, and ecological objectives. Four individual value criteria are considered to be relevant, assessed by two types of benefits and of costs, respectively. Benefit $b_1$ represents the interests of the shareholders of the DMUs and benefit $b_2$ the social interest for employment. The two relevant types of costs $c_1$ and $c_2$ represent ecological objectives and are concerned with the contributions of the DMUs towards climate change through global warming and to the ozone hole in the stratosphere, respectively.

It is necessary to identify which types of inputs and outputs as well as which types of impacts are relevant insofar as they determine the considered costs and benefits. Let us assume that the responsible evaluator establishes the following impacts and corresponding impact-related value functions:

$$b_1 = r_\pi = r_{rev} - r_{exp} \qquad c_1 = r_{RFI} = r_{CO2} + 8500 r_{CFC}$$
$$b_2 = r_{emp} \qquad c_2 = r_{OZN}$$

as well as the following six relevant types of input and output and corresponding impact functions:

$$r_{rev} = 40y_1(10 - y_1) + 20x_3 \qquad r_{CO2} = y_2 \qquad r_{emp} = x_1$$
$$r_{exp} = 10x_1 + 50x_2 \qquad r_{CFC} = y_3 \qquad r_{OZN} = y_3$$

Impact assessment and valuation imply the respective four process-related value functions:

$$b_1 = 40y_1(10 - y_1) - 10x_1 - 50x_2 + 20x_3 \qquad c_1 = y_2 + 8500y_3$$
$$b_2 = x_1 \qquad c_2 = y_3$$

For example, the DMUs may be cement plants with labour $x_1$, raw material $x_2$ and scrap tires $x_3$ as inputs as well as cement $y_1$, carbon dioxide $y_2$ ($CO_2$) and chlorofluorocarbon $y_3$ (CFC) as outputs. The shareholders' subjective benefit is specified by the profit resulting from financial transactions on markets as net total from revenues and expenditures. Here, the DMUs represent distinct business units selling the produced cement on their respective local markets, each of which is determined by a linear demand function ($y_1 = 10 - p/40$) with respect to the individual price $p$ per quantity unit set by each DMU. Financial turnover does not only result from sales of the cement but also from revenues generated by the factory imposing a fee on the disposal of used tires. These scrap tires are incinerated as an input, thereby serving as fuel for the process of cement production. Expenses for wages and salaries as well as those for raw material purchases add up to the total expenditures. The second benefit, job creation, can be gauged from employment figures of labour input, and the greenhouse effect is calculated as radiative forcing (RFI) based on emissions of carbon dioxide ($CO_2$) and chlorofluorocarbon (CFC), the latter of which also causes damage to the ozone layer (OZN).

In the case of four cement plants with following matrices $\boldsymbol{X}$ and $\boldsymbol{Y}$ displaying their input and output quantities, the matrices $\boldsymbol{C}$ and $\boldsymbol{B}$ show the respective costs and benefits:

Production activities

$$\boldsymbol{X} = \begin{array}{cl} \begin{bmatrix} 4 & 4 & 5 & 3 \\ 3 & 5 & 5 & 5 \\ 5 & 1 & 3 & 3 \end{bmatrix} & \begin{array}{l} \text{labour} \\ \text{raw material} \\ \text{scrap tires} \end{array} \end{array}$$

$$\boldsymbol{Y} = \begin{array}{cl} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 120 & 40 & 100 & 100 \\ 0.6 & 0.2 & 0.5 & 0.5 \end{bmatrix} & \begin{array}{l} \text{cement} \\ CO_2 \\ \text{CFC} \end{array} \end{array}$$

$$\Longrightarrow$$

Created and destroyed values

$$\boldsymbol{B} = \begin{array}{cl} \begin{bmatrix} 270 & 90 & 120 & 140 \\ 4 & 4 & 5 & 3 \end{bmatrix} & \begin{array}{l} \text{profit} \\ \text{employment} \end{array} \end{array}$$

$$\boldsymbol{C} = \begin{array}{cl} \begin{bmatrix} 5220 & 1740 & 4350 & 4350 \\ 0.6 & 0.2 & 0.5 & 0.5 \end{bmatrix} & \begin{array}{l} \text{climate effect} \\ \text{ozone effect} \end{array} \end{array}$$

The example demonstrates three aspects which are unusual for traditional economics and performance evaluation. Firstly, labour input $x_1$ has two opposing impacts, an undesired financial impact on profit $b_1$ and a desired social one on employment $b_2$. Thus, the fourth DMU is economically better off than the third, but worse off from a social perspective. Secondly, the output CFC $y_3$ has two different, undesired ecological impacts at the same time. Thirdly, scrap tires are considered here as bads (undesirable objects) whose input $x_3$ into incineration is thus desired as it destroys them and adds value by reducing negative values.

In a further step of aggregation, well-known methods of MCDA may be applied to determine a top KPI $u = u(\boldsymbol{v}) = u(c_1, c_2; b_1, b_2)$ that represents the overall preferences which are decisive for the evaluation. Literature on various concepts and approaches used for such a setting and related purposes is reviewed and discussed in Section 4. With reference to the example, Section 5 explains some general theoretical propositions which have recently been published and are relevant for these purposes. They are based on the following combination of production and decision theory.

### 2.4. Basics of multicriteria production theory

According to the basic structure of decision models mentioned in Section 2.1.2, five fundamental assumptions characterise the corresponding instance of a general *multicriteria production theory* (MCPT). It extends the one introduced by Dyckhoff (2018) in such a way that it fully fits with the framework of Section 2.2.[3]

**A1**: The activities of PPS $\mathcal{P}$ are completely described by a vector $\boldsymbol{z} = (\boldsymbol{x}; \boldsymbol{y})$ of $m$ input and $s$ output quantities of certain selected types of objects involved in the transformation process. Basically, $\mathcal{P}$ is part of a technology $\mathcal{T}$ which is defined by certain axioms (e.g. free disposal) and individual characteristics (e.g. constant returns to scale) as predetermined general or specific properties:
$\mathcal{P} \subset \mathcal{T} = \{\boldsymbol{z} = (\boldsymbol{x}; \boldsymbol{y}) \in \mathbb{R}_+^{m+s} |$ Input $\boldsymbol{x}$ can in principle be transformed into output $\boldsymbol{y}\}$
$\mathcal{P}$ is often generated by some basic activities and restricted by constraints given for the decision at hand.
**A2**: There is no uncertainty in the data.
**A3**: Relevant consequences of any production activity are completely captured by a multi-dimensional impact function $\boldsymbol{r}(\boldsymbol{z}) \in \mathbb{R}^q$ of the respective input/output-vector $\boldsymbol{z} = (\boldsymbol{x}; \boldsymbol{y})$ that distinguishes all relevant results caused by the inputs and outputs of the transformation process. The image $\boldsymbol{r}(\mathcal{P})$ of the PPS is called *impact possibility set* (IPS).
**A4**: Relevant evaluation criteria are measured by a multi-dimensional value function $\boldsymbol{v} = \boldsymbol{w}(\boldsymbol{r}) \in \mathbb{R}^{k+\ell}$ of the impacts $\boldsymbol{r} = \boldsymbol{r}(\boldsymbol{z})$ and are differentiated into two distinct (usually non-negative) categories $\boldsymbol{v} = (\boldsymbol{c}; \boldsymbol{b})$, namely $k$ types of values destroyed, called *costs*, representing disadvantageous results $\boldsymbol{c}$, as well as $\ell$ types of created values as advantageous results $\boldsymbol{b}$, that are called *benefits*. Objectives are both the minimisation of each type of cost as well as the maximisation of each type of benefit. The image $\boldsymbol{v}(\mathcal{P}) = \boldsymbol{w}(\boldsymbol{r}(\mathcal{P}))$ of the PPS is called *value possibility set* (VPS).
**A5**: The preferences of the responsible decision-maker or evaluator are compatible with the vector dominance relations of the alternatives regarding the values in the $(k+\ell)$-dimensional space of costs and benefits.

Thus, MCPT is determined by the premise that all relevant data are known, deterministic and measurable. Apart from the above more structural assumptions, $\mathcal{P}$ and $\mathcal{T}$ should regularly be *closed* and *non-trivial*, i.e. contain at least two different activities, whereas the multiple functions $\boldsymbol{r}(\boldsymbol{z})$ and $\boldsymbol{w}(\boldsymbol{r})$ should be *continuous* (and will be non-linear in general). Since $\mathcal{P}$ is defined as that part of the technology $\mathcal{T}$ which is realisable in the situation at hand, it is furthermore *bounded* in practice, e.g. by resource restrictions; then, the IPS $\boldsymbol{r}(\mathcal{P})$ and the VPS $\boldsymbol{v}(\mathcal{P})$ are closed and bounded, too.[4]

Definition: *Efficiency and Effectivity*

(a) An activity is *(strongly) efficient* with respect to $\mathcal{P}$ and the multiple costs and benefits as relevant objectives if there is no other alternative in $\mathcal{P}$ dominating it. Here, an activity *dominates* another one (*weakly*) if it is better for at least one objective and not worse regarding all others.
(b) An activity is *weakly efficient* if it is not strongly dominated in regard of all relevant objectives and all other possible activities. Here, an activity *dominates* another one *strongly* if it is truly better for each of the relevant objectives.
(c) An activity is *effective* if it is efficient with regard to the ends, i.e. not dominated with respect to those values that originally motivate the activity (thus excluding means and secondary results).

All traditional theories of production and cost fulfil assumptions A1–A5, but specify them by additional, very particular requirements concerning the value function and, possibly, also the PPS. With respect to $\boldsymbol{v}(\boldsymbol{x}; \boldsymbol{y}) = \boldsymbol{w}(\boldsymbol{r}(\boldsymbol{x}; \boldsymbol{y}))$, only the following two extreme cases of a continuum of specialised theories are explored.

At the one extreme, there exists only a single one-dimensional value function to be maximised, i.e. $v(\boldsymbol{x}; \boldsymbol{y}) \in \mathbb{R}$. It measures the success as overall benefit generated by the inputs and outputs of the production process and is usually determined by the profit or contribution margin (such as by formula (1) in Section 3.1). Here, assumption A4 reduces itself to the trivial case: $\ell = 1$, $k = 0$, so that $v(\boldsymbol{x}; \boldsymbol{y}) = b(\boldsymbol{x}; \boldsymbol{y})$. If the revenues are supposed to be fixed, *traditional cost theories* are obtained, i.e. $k = 1$, $\ell = 0$ with $v(\boldsymbol{x}; \boldsymbol{y}) = c(\boldsymbol{x}; \boldsymbol{y})$. In these one-dimensional cases, the above concept of efficiency is simplified to 'greatest success', i.e. maximum profit or minimum cost, respectively.

At the other extreme, *traditional production theories* alternatively consider the simplest case of what may constitute the relevant consequences and values of a production activity. Each selected type of objects that are involved in the production process uniquely forms one of the $k = m$ types of costs on the input side: $\boldsymbol{c}(\boldsymbol{x}; \boldsymbol{y}) = \boldsymbol{x}$, or one of the $\ell = s$ types of benefits on the output side: $\boldsymbol{b}(\boldsymbol{x}; \boldsymbol{y}) = \boldsymbol{y}$ (cf. formula (4) in Section 3.3). In this case, the above definition of efficiency is reduced to the usual one well-known from Koopmans (1951), also called *technical efficiency* (Farrell, 1957).

Other special cases of MCPT can be achieved by different specifications of the multi-dimensional impact and value functions $\boldsymbol{v}(\boldsymbol{x}; \boldsymbol{y})) = \boldsymbol{w}(\boldsymbol{r}(\boldsymbol{x}; \boldsymbol{y}))$, as well as of the fundamental axioms and particular properties by further specifying the technology $\mathcal{T}$ and the PPS $\mathcal{P}$. For example, an *ecological production theory* may use an impact function (like formula (2) in Section 3.2) which measures the global warming impact of various greenhouse gases in terms of carbon dioxide equivalents (Dyckhoff & Allen, 2001).

## 3. Specific production-based methodologies of performance evaluation

Distinct (not only production-based) methodologies for performance evaluation may be classified into important categories according to the specific set of evaluated actions as well as to the strength of evaluation. In practice, the evaluation set $\mathcal{E}$ nearly always consists of a finite, manageable number of actions the (input and output) relevant results of which have been observed in the past. In extreme cases, $\mathcal{E}$ consists of a single action, only ($|\mathcal{E}| = 1$).

[3] Assumption A3 of Dyckhoff (2018) does not explicitly address the impacts such that costs and benefits are directly defined on the quantities of the inputs and outputs.

[4] Furthermore, in economics and DEA, it is usually assumed that the PPS is convex or even (non-negatively) linear. However, it must be noted that a linear PPS is not bounded. The assumption of linearity presupposes empirically that the target points and benchmarks found mathematically by a linear data envelopment of observed activities lie within the restrictions given in practice. In a similar vein, the assumption of a linear or convex PPS contradicts reality if some of the inputs or outputs exist in integer quantities, only. Nevertheless, if these quantities are large enough, the rounding error in the performance results may often be acceptable.

Then, in order to decide whether the action has been successful regarding a specific objective, the evaluation requires either an absolute scale determining the degree of success or, alternatively, some externally set *benchmarks* or *standards* with which it can be compared.

Such types of absolute evaluations contrast with *relative* ones which compare several actions with each other and do not require exogenously determined benchmarks or standards. Two basic types can be distinguished: either the set $\mathcal{A}$ of actually compared actions equals the evaluation set ($\mathcal{E} = \mathcal{A}$) or $\mathcal{E}$ is *enlarged by additional comparable actions* ($\mathcal{E} \subset \mathcal{A}$). Cases in which additional suitable actions are given externally resemble the benchmark or standard comparisons mentioned before. Other types of methodologies use some general, often technological, external information, determined a priori, to construct comparable actions within the performance analysis which might also have been realised in addition to the observed ones of the evaluation set $\mathcal{E}$. Furthermore, we distinguish between methods in which the set $\mathcal{A}$ of compared alternatives is either *finite or infinite* (analogously to the well-known differentiation in MCDM).

It is characteristic for production-based methodologies that extensions $\mathcal{A} \supset \mathcal{E}$ of the evaluation set $\mathcal{E}$ are substantially determined by the PPS to which the observed actions belong, i.e. $\mathcal{A} = \mathcal{P}$ or $\mathcal{A} \subset \mathcal{P}$. In general, a production-theoretically motivated and justified *data envelopment* of the evaluation set $\mathcal{E}$ is achieved by the smallest enlargement or 'minimal extrapolation' so that $\mathcal{A}$ abides by the same technological rules and possibly also adheres to the restrictions that define the PPS: $\mathcal{E} \subset \mathcal{A} \subset \mathcal{P}$. However, the higher the number of possible actions with good performance in set $\mathcal{A}$, the stricter the assessment may be.

After having made the distinction of several basic types of performance analyses clear, we will now illustrate them by elaborating on the three well-known methodologies mentioned in the introduction. All concepts and methods of production-based performance evaluation are characterised by the fact that they assess exactly those actions $a \in \mathcal{E}$ which are essentially represented by the quantities $(\boldsymbol{x}; \boldsymbol{y})$ of inputs entering and outputs emerging from a transformation process which are in turn largely identified with the action $a$ itself.

### 3.1. Financial and management accounting

In economics, a single production process is classified as absolutely successful if it is profitable, irrespective of other possible activities. The goods and services representing either input or output are valued at the market prices $\boldsymbol{\eta}$ and $\boldsymbol{\mu}$ at which they are purchased or sold, as a rule. The expenditures $r_{exp} = \boldsymbol{\eta x}$ for the input quantities $\boldsymbol{x}$ as well as the revenues $r_{rev} = \boldsymbol{\mu y}$ from the output quantities $\boldsymbol{y}$ form two kinds of impact functions which are directly comparable because of the underlying financial transactions. By using the same monetary scale, the expenditures can be subtracted from the revenues which results in the profit or contribution margin $r_\pi$ that is a desired outcome, i.e. it represents a benefit

$$b(\boldsymbol{x}; \boldsymbol{y}) = r_\pi(\boldsymbol{x}; \boldsymbol{y}) = r_{rev} - r_{exp} = \boldsymbol{\mu y} - \boldsymbol{\eta x} \tag{1}$$

Because of quantity-dependent prices $\boldsymbol{\eta} = \boldsymbol{\eta}(\boldsymbol{x})$ and $\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{y})$, the numerical example of Section 2.3 has illustrated that impact functions for expenditures, revenues and profit may also be non-linear, e.g. in cases of monopolists or all-units quantity discounts. If market prices are not immediately obvious, it is a main task of financial and management accounting to find other monetary valuations $\boldsymbol{\eta}$ and $\boldsymbol{\mu}$ for the inputs and outputs such that $b(\boldsymbol{x}; \boldsymbol{y}) = \boldsymbol{\mu y} - \boldsymbol{\eta x}$ still forms a single top KPI reflecting the objectives of the shareholders. Usually based on trade-offs which are derived from existing markets, appropriate valuations should allow a direct comparability of the benefits and costs induced by the inputs and outputs.[5] Regarding the generic structure of performance analysis, shown in Figure 2, this usually entails some (perhaps implicit) steps of impact assessment.

Contrary to the valuations of goods as inputs or outputs, the valuations of bads are negative. Therefore, costs – as consequences of a production activity that you wish to reduce – result not only from the input of goods (because of expenditures for buying them), but also from the output of bads (because of emission fees to be paid). On the opposite, benefits – as consequences to be maximised – either result from the output of goods (usually due to the revenues from selling them) or from the input of bads (e.g. because an incineration plant collects an acceptance fee for the waste it burns).

### 3.2. Life cycle assessment

Regarding the eco-efficiency analysis of BASF as a particular tool box for LCA following ISO 14040 and 14044 (Grosse-Sommer et al. (2020)), Figure 1 has already illustrated that there may be several successive environmental valuation steps, instead of the single one in the generic picture of Figure 2.[6] BASF uses linear functions in each valuation step, the weighting coefficients thus determining a constant rate of preference trade-off between the different ecological impacts. A given set $\mathcal{E}$ of products or technologies is compared with each other ($\mathcal{A} = \mathcal{E}$).[7] The assessment results in two top KPIs: the overall environmental impact and the monetary life cycle costs, each measured in relation to the arithmetic mean of all alternatives evaluated.

An intersubjectively reproducible assessment of products or technologies may be conducted in respect of some definitive legal requirements or licence specifications in the context of permission for use. The concrete valuation of the trade-off between different ecological impacts is often exogenously established by experts and derived from given emission limits or standards. Generally, an objective valuation of ecological impacts regarding the fundamental objectives is very difficult if not impossible.

This cardinal problem of valuation is illustrated by *Figure 3* for the case of the greenhouse effect as a severe ecological impact that threatens the existence of civilised societies in future. Global warming is based on a complex network of impacts of at least four types of gaseous emissions. Some of these impacts are very long-term and difficult to assess. Figure 3 shows only an excerpt of the manifold, serious consequences for society and nature. In addition to the effects of flooding and drought mentioned, there is a much greater number of other highly plausible direct and indirect impacts of atmospheric temperature rise, which, according to current medical knowledge, lead to severe impairment of the quality of human life (Eis et al. (2010)).

Therefore, LCA methods usually do not try to model and to assess the various (moreover widely unknown) consequences with respect to the fundamental goals as endpoints of the network of

[5] Since appropriate markets where comparable goods and services are traded often do not exist, other valuation methods are necessary, too. In such cases as these, management accounting applies well-known concepts from decision theory and in general defines benefits and costs as the advantages and disadvantages, respectively, induced by the chosen action (analogously to Sections 2.1.1 and 2.1.2; cf. Ewert & Wagenhofer (2014), p. 35).

[6] Figure 2 is similar to visualisations of the LCA framework by ISO 14040. There, however, the valuation is usually part of the interpretation and often called "evaluation". Occasionally, the importance of valuation for LCA is emphasised analogously to our framework by including it as a separate step, e.g. by Thies et al. (2019).

[7] Since the weighting coefficients partly employ *fixed* factors for social preferences, the valuation is subject of the so-called 'range effect'. Therefore, BASF's eco-efficiency method in its original version violates the rationality principle of *independence from irrelevant alternatives* (Dyckhoff, Quandel, & Waletzke, 2015). The specific choice of the evaluation set $\mathcal{E}$ may thus be of decisive importance.

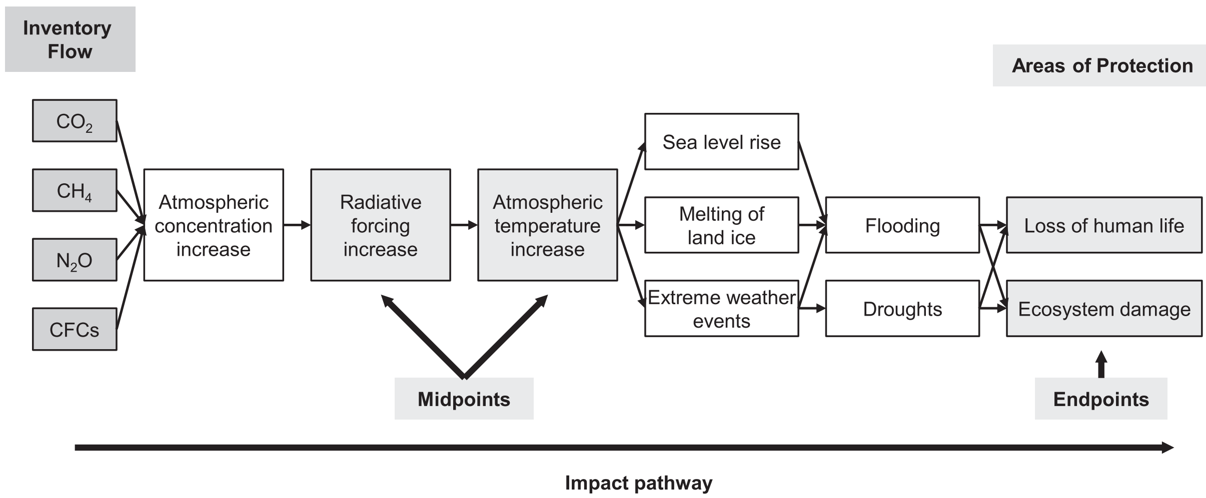


**Fig. 3.** From outputs via impacts to fundamental goals (cf. Hauschild & Huijbregts (2015), Fig. 1.2)

impacts themselves, such as the loss of human life or of biodiversity. Instead, they try to indirectly take them into account by directly valuing certain midpoints which often can be assessed with a higher degree of validity, owing to objective knowledge about laws of nature. In Figure 2, such midpoints build the bridge between the impact assessment and the valuation steps.

Regarding global warming in Figure 3, for example, the increase of radiative forcing (RFI) or the rise of atmospheric temperature are such midpoints. This allows for an aggregation of the different impacts of greenhouse gases. To measure their global warming potential (GWP) as environmental cost $c_{\mathrm{GWP}}$, a typical result function $r_{RFI}(\boldsymbol{x};\boldsymbol{y})$ comprising the total impact of emissions of carbon dioxide ($CO_2$), methane ($CH_4$), nitrous oxide ($N_2O$), and chlorofluorocarbons (CFC) reads as follows (cf. the example of Section 2.3):

$$c_{\mathrm{GWP}} = r_{RFI} = y_{\mathrm{CO2}} + 28y_{\mathrm{CH4}} + 265y_{\mathrm{N2O}} + 8500y_{\mathrm{CFC}} \tag{2}$$

This way, the negative climate effects of the greenhouse gases are summed up by means of so-called carbon dioxide equivalents. Here, one mass unit of methane has about 28 times more greenhouse potential over 100 years than one mass unit of carbon dioxide, and that of the other two greenhouse gases is accordingly even higher (Myrhe & al. (2013), p. 731). Choosing a shorter time horizon as period of evaluation implies that methane has an even much higher impact (more than 28 times) whereas most methane will have left the atmosphere within more than thousand years or will have been converted into carbon dioxide that stays in the atmosphere for long. These facts disclose that already the step of impact assessment in LCA necessarily needs some value decisions, in this case concerning the choice of the relevant time horizon (Lueddeckens, Saling, & Guenther, 2020).

### *3.3. Data envelopment analysis*

Whereas in LCA the set $\mathcal{A}$ of compared alternatives usually equals the finite evaluation set $\mathcal{E}$, DEA is characterised by an extension of the evaluation set $\mathcal{E} = \{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n\}$ where $\boldsymbol{z}_j = (\boldsymbol{x}_j;\boldsymbol{y}_j) \in \mathcal{E}$ describes the observed action of DMU $j$. $\mathcal{E}$ is enveloped by an infinite number of comparable alternatives, to be justified by information about the actual PPS $\mathcal{P}$ which is available beforehand so that $\mathcal{E} \subset \mathcal{A} \subset \mathcal{P}$. For example, if the technology is linear, the associated smallest enlargement of $\mathcal{E}$ becomes:

$$\mathcal{A} = \left\{ \boldsymbol{z} = \sum_{j=1}^{n} \lambda_j \boldsymbol{z}_j \middle| \boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n) \geq \boldsymbol{0} \right\} \tag{3}$$

On the contrary, and again in stark contrast to LCA, the original conception of DEA is characterised by the fundamental premise that no further information about impacts and values is known which would allow for an aggregation of the considered inputs and outputs, except for the following general assumption about the preferences: Output is desirable and input undesirable! Thus, regarding the generic framework of production-based performance evaluation shown in Figure 2, there is no step of impact assessment, and the valuation step is reduced to the trivial case where multiple benefits are measured by the physical quantities of the relevant outputs and the costs by those of the inputs:

$$\boldsymbol{b}(\boldsymbol{x};\boldsymbol{y}) = \boldsymbol{y} \quad \text{and} \quad \boldsymbol{c}(\boldsymbol{x};\boldsymbol{y}) = \boldsymbol{x} \tag{4}$$

For instance, in order to measure the efficiency of an action $\boldsymbol{z}_o = (\boldsymbol{x}_o;\boldsymbol{y}_o) \in \mathcal{E}$ in relation to the linear envelopment (3), the input-oriented *CCR model* in envelopment form – building one of the pioneering models of Charnes, Cooper & Rhodes (1978) – is given by following linear programme (LP):[8]

$$\theta_o^* = \min_{\boldsymbol{\lambda} \geq \boldsymbol{0}} \theta_o \text{ such that } \sum_{j=1}^{n} \lambda_j \boldsymbol{x}_j \leq \theta_o \boldsymbol{x}_o \text{ and } \sum_{j=1}^{n} \lambda_j \boldsymbol{y}_j \geq \boldsymbol{y}_o \tag{5}$$

## 4. Literature linking production with multiple criteria for performance analysis

Thus, DEA, LCA, and accounting are methodologies, well-known from various scientific disciplines and applied in practice, which all fit in a specific, but diverse manner as particular cases into the framework presented in Section 2. All three evaluate the performance of activities that transform several inputs into several outputs. However, none of them fully integrates multicriteria analysis with production theory for that purpose: Whereas business accounting as well as DEA are indeed based on concepts of production theory, particularly the PPS and its properties, like e.g. convexity or returns to scale, multiple criteria are not at all considered in traditional accounting, and they are only implicitly identified with inputs and outputs as trivial cases of multiple objectives in common DEA. In contrast, LCA is usually not based on production theory. However, since LCA considers ecological impacts as

[8] Radial DEA models like (5) are usually complemented by an infinitesimally small summand in the objective function or by a second optimisation step identifying possible slacks for individual inputs or outputs to avoid weak efficiency. For reasons of simplicity, we neglect such amendments here, because they do not affect our main reasoning.

multiple objectives, a growing number of articles which explicitly use MCDA methods in environmental sciences has been published since the 1990s.[9] Yet, reviews by Martin-Gamboa et al. (2017), Campos-Guzman et al. (2019) and Thies et al. (2019) show that this takes place mostly in scientific journals of application areas not focused on *Operations Research and Management Science* (OR/MS).

The significant differences between the three basic types of production-based methodologies of performance evaluation, characterised in Section 3, lead to the question whether the existing literature provides encompassing, theoretically well-grounded concepts and methods of performance evaluation that systematically integrate the production with the multicriteria perspective. If this should not be the case, how much do existing approaches linking both perspectives fit in with this claim and with the generic framework of production-based performance evaluation developed in Section 2? To answer these questions, a systematic narrative literature review has been conducted.[10] Our purpose is to identify – in a transparent and reproducible way – relevant ideas, concepts, and approaches as well as the corresponding main literature that has developed, discussed, and applied them. In that sense, we focus on representative sources, and not on a comprehensive bibliography of all contributions which have been further developing specific aspects or only applying these ideas, concepts, and approaches without major conceptual improvements.

The next subsection describes our search process. Representative results are collected and then categorised in Section 4.2. Section 4.3 discusses this literature as well as existing gaps, thus yielding main arguments for our thoughts on possible future research paths in the concluding Section 6.

### 4.1. Search for relevant literature

To find relevant literature we adopted a four-step approach which is shown by *Figure 4*. The search was conducted mainly within the OR/MS category of the *Web of Science* core collection database and included articles published until the end of 2019. The schematic analysis via the different search strings[11] presented in Figure 4 refers to the title, abstract and keywords of the articles. The screening process assessing the relevance also included a closer review of the whole article where it seemed useful. The four-step approach started from a core search regarding papers concerning all three relevant topics (A) *Performance Evaluation*, (B) *MCDA* and (C) *Production*. Due to the small number of sources found, this – perhaps too narrow – search was extended by further search strings replacing the production topic by the three performance methodologies (D) *Accounting*, (E) *LCA* and (F) *DEA*. In addition, relevant reviews concerning topics (A) and (B) were focused on and the database was extended accordingly.

Step 1: Although the search for the main topics (A), (B) and (C) leads to a huge number of sources for each single topic,[12] less than hundred papers deal with a combination of all three topics. The most frequently cited of these sources is by Dyckhoff & Allen (2001). It belongs to those 20 of them which fall into OR/MS category. There is only one review – by Soheilirad et al. (2018) – that refers to all three topics simultaneously. However, its focus is on DEA with merely some short comments on MCDA. The present review intends to fill in this obvious gap. A cursory look is sufficient to notice that at least nine of the 20 OR/MS sources found are not at all concerned with the specific methodological aspects of performance analysis that are of interest for our review. Consequently, the residual eleven sources constitute the *core literature* to be further analysed and categorised in Section 4.2.

Step 2: In order to find more promising sources on performance evaluation that eventually link the production with the multicriteria perspective, we searched for reviews of literature applying MCDA methods to the performance evaluation of processes transforming inputs into outputs.[13] We found 36 reviews, eight of which (including the review by Soheilirad et al. (2018) already found in Step 1) are useful to gain more insights. From their references thirteen further sources have been selected as representative and have been added to the literature to be analysed in Section 4.2.

Step 3: Although – or rather due to the fact that – the specific performance evaluation methodologies discussed in Section 3 are production-based, papers referring to them often do not mention any terms of the production topic (C) in their title, abstract and keywords. Thus, we extended the prior steps by a search which combines (A) and (B) with one of the topics (D), (E) or (F).[14] By focusing our screening on those articles and corresponding ideas, concepts, or approaches which are not already captured by the reviews found in Step 2, further 19 sources have been selected.[15]

Step 4: To extend the search area further we looked for sources not listed in the Web of Science. In doing so, we used the core literature of eleven articles found in Step 1 as starting point. At first, a backward citation analysis identified all the other (earlier) sources referenced by them and not already found before – notably papers in books, proceedings and journals not documented in the OR/MS category of the Web of Science. After a deeper screening, only one additional source has been added to the relevant literature. Starting again with the core literature of Step 1, a forward citation analysis by *Google Scholar* identified later sources up to the year 2019. In this way, nine further sources have been selected for categorisation and discussion.

At the end, the overall list for a more thorough investigation contains 60 sources.[16] Because of the selections made during the four-step search process it does not capture the whole relevant literature. Nevertheless, the list can be seen as representative with respect to the principal OR/MS ideas, concepts and approaches of

[9] Among others Bloemhof-Ruward, Koudijs & Vis (1995), Miettinen & Hämäläinen (1997) and Azapagic & Clift (1999).

[10] Thus, we follow suggestions of Fisch & Block (2018) – in contrast to those for bibliographic studies (Block & Fisch, 2020). Only English language sources have been considered for this investigation.

[11] The sign * within the search strings indicates that any letter may lead or follow the given *string of letters* in order not to miss sources with a slightly different wording.

[12] Although more than 5.2 million sources are concerned with performance topic (A), only 50,024 belong to the category OR/MS, amongst them 766 reviews. The by far most frequently cited OR/MS source is the pioneering article of Charnes, Cooper & Rhodes (1978) that introduced DEA as a production-based method of performance measurement. From a total of about 120,000 sources regarding the MCDA topic (B), 12,922 are in the OR/MS category, with 253 of them being reviews. The most frequently cited of these reviews are concerned with AHP, rough sets, TOPSIS, or PROMETHEE as particular methods or concepts for MCDA and with applications in the field of supplier evaluation and selection. The most frequently cited OR/MS source ever dealing with topic (B) is the paper of Saaty (1990) on AHP. In comparison, the production topic (C) leads to far fewer sources, i.e. about 21,000 in total, with 1,005 of them in the OR/MS category, amongst the latter 15 reviews. Three of the five most often cited reviews deal with DEA, namely Dakpo, Jeanneaux & Latruffe (2016), Olesen & Petersen (2016), and Soheilirad et al. (2018). Additionally, nine of the twenty OR/MS sources which are concerned with topic (C) and record the highest citation counts also mention DEA in their titles explicitly.

[13] The Web of Science lists a total of about 34,000 entries combining topics (A) and (B), 3,564 of which are in OR/MS category.

[14] The search in the Web of Science for each of the three topics (D), (E), and (F) alone yields about 193,000, 411,000, and 28,000 sources in total of which 2,148, 3,035, and 4,022 are in OR/MS category, with 51, 68, and 69 being reviews.

[15] OR/MS sources combining MCDA with *financial* accounting are notably seldom; Sueyoshi, Shang & Chiang (2009) and Yalzin, Bayrakdaroglu & Kahraman (2012) form such rare examples, but they are not selected in our search process as they do not provide any relevant idea, concept or approach to our subject of investigation.

[16] Table A1 of Appendix A lists these sources and assigns them to the category of its main concern (defined in Section 4.2).

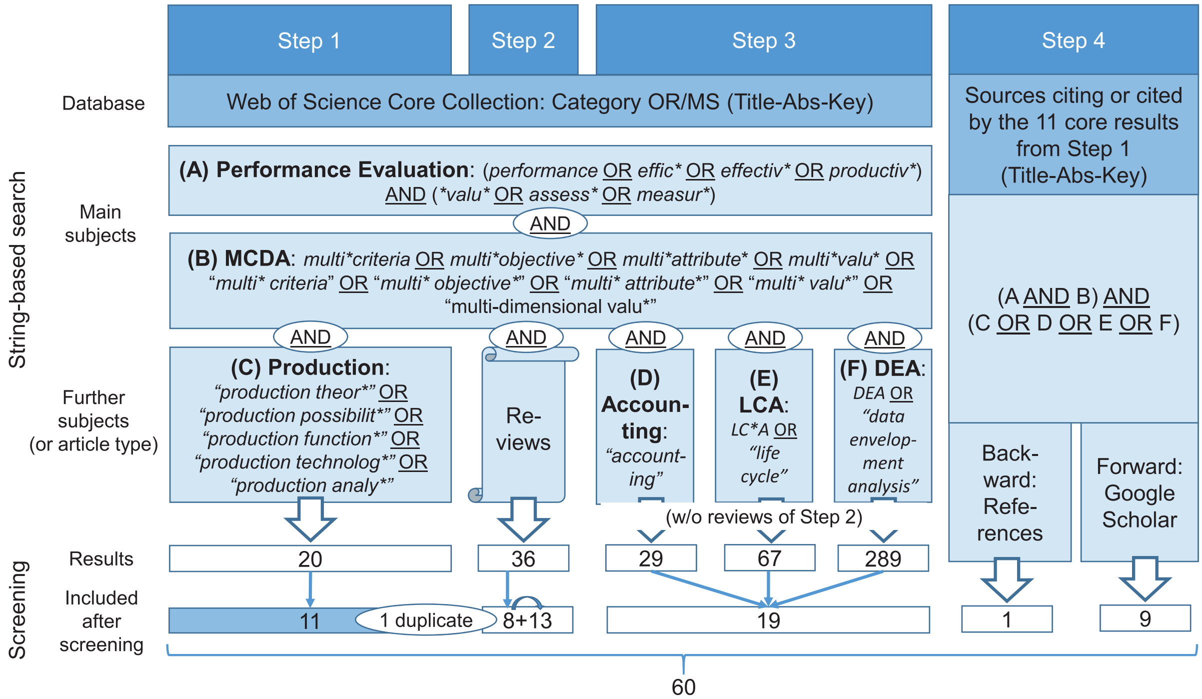


**Fig. 4.** Search process for relevant literature

performance evaluation linking the production with the multicriteria perspective.

### 4.2. Categorisation of relevant ideas, concepts, and approaches

In fact, there exist tens of thousands of scientific articles for each single search topic from (A) to (F). Nonetheless, OR/MS literature combining multicriteria performance evaluation (topics A and B) with production theory (C) or with the production-based topics of accounting (D), LCA (E), or DEA (F) is scarce. Based on the framework and types developed in Sections 2 and 3, several principal categories of concepts and approaches appearing in this literature are now distinguished and further subdivided if it makes sense with respect to our purposes. This differentiation is rather a kind of typification than a strict classification so that the categories may overlap. Each of the categories (or types) will be explained and exemplified by selected representative (typical) sources found by our search. Furthermore, already existing reviews of the associated literature as well as pioneering articles are mentioned and utilised, thereby incorporating further information that we gained while compiling and studying the literature during the search process explained before.

#### 4.2.1. Multicriteria evaluation of a fixed finite set of production activities (Category 1)

The literature of the first main category restricts the evaluation to a given finite set $\mathcal{E}$ of activities or objects to be assessed ($\mathcal{A} = \mathcal{E}$). It is already well described by several recent reviews, most of which found in Step 2 of our search process (Brandenburg et al. (2014), Eskandarpour et al. (2015), Ilgin, Gupta & Battaia (2015), Banasik et al. (2018), Thies et al. (2019)). Their reviewed sources are usually concerned with a specific real-world domain of performance evaluation. They directly apply multicriteria methods to manifold production impacts as attributes of defined objectives, particularly to those of various manufacturing, logistics, and energy producing systems as well as whole supply chains – and even beyond. A wide range deals with the measurement of ecological and sustainability performance (e.g., Ng & Chuah (2014) use AHP in LCA). As a rule, the multicriteria methods are applied to the impacts of the products or production activities to be evaluated, but do not refer to the quantitative relationships of the inputs and outputs evoking these impacts.[17]

#### 4.2.2. Weak links between MCDA and production theory (Category 2)

Most OR/MS literature found in Section 4.1 is concerned with methodologies which enlarge the given evaluation set $\mathcal{E}$ by additional fictitious comparable alternatives to a set $\mathcal{A}$ (i.e. $\mathcal{A} \supset \mathcal{E}$), ideally by assuming certain technological properties of production possibilities so that $\mathcal{A} \subset \mathcal{P}$. Depending on the strength of the link between both perspectives of performance evaluation, three additional main categories can be distinguished, namely (2) *weak*, (3) *strong*, and (4) *full* integration of MCDA and production theory. As we will see, there exists no literature for the fourth category, i.e. which systematically elaborates sophisticated methods of production-based multicriteria performance evaluation that are appropriately applicable to examples such as the one of Section 2.3.

The second main category encompasses those sources which utilise or discuss merely weak links between MCDA and production-based performance evaluation. It is divided into two subcategories.

Category 2a: *Moderately combining MCDA methods and production models in real-world applications*

Besides the approaches of Category 1, the reviews mentioned there also capture similar ones that explicitly consider an infinite instead of a finite number of production possibilities. However, as a rule, they do not apply elaborated MCDA methods. Usually, such approaches develop methodologies that assess the *eco-efficiency* of production activities, logistics networks or supply chains by analysing the trade-off between an economic objective on the one

[17] The paper of Parkan & Wu (1997) forms an exception. Analogously to Category 2b, but with respect to a finite number of possible alternatives only, it examines the equivalence of operational performance measurement and multiple attribute decision making (MADM). In particular, the authors compare specific tools, namely OCRA (Operational Competiveness Rating) with TOPSIS (Technic for Order Preference by Similarity to Ideal Solution).

hand and one or more ecological objectives on the other hand. For that purpose, they typically use bi- or tri-objective (mixed-integer) linear programming models to calculate the respective efficient frontier (cf. Quariguasi Frota Neto et al. (2009) for a literature review and a specific methodology).

Some of the reviews found in Step 2 also include articles dealing with DEA (Ho, Xu & Dey (2010), Ilgin, Gupta & Battaia (2015), Thies et al. (2019)). By referring to Cooper, Seiford & Tone (2007), Martin-Gamboa et al. (2017) state as reason for choosing DEA as a main performance measurement tool in real-world contexts: "Within the MCDA tools ... DEA arises as a trade-off solution between soundness and practicality" (p. 179). Nevertheless, for a more comprehensive performance evaluation, DEA has increasingly been combined with other methods that are more specific for MCDA, e.g. with AHP/ANP by Yang & Kuo (2003) or with ELECTRE TRI by Madlener, Antunes & Dias (2009), whereas in all cases DEA is applied to a particular evaluation topic.[18] However, such combinations of DEA with one of the various well-known MCDA methods are usually chosen *ad hoc*. The contributions lack a closer analysis of the production context, and the MCDA method is typically included without any theoretical reasoning or empirical validation.

The marginal consideration of production relations seems to be typical for a relatively new research path, too. It integrates LCA with DEA structurally in approaches consisting of several steps (e.g. Lozano et al. (2009), Iribarren et al. (2010), Vazquez-Rowe et al. (2010)). This literature intends to apply DEA as a multicriteria method either directly to the inputs and outputs determined by the life cycle inventory or to their impacts resulting from it (corresponding to the second or third step of the generic structure of performance evaluation in Figure 2). In this way, a basis for a conceptual integration of multicriteria and production-based life cycle assessment is laid. In their review, Martin-Gamboa et al. (2017) justify this integration of LCA and DEA by the fact that it "benefits from the advantages of both methodologies while overcoming some of their limitations" (p. 171).

Category 2b: *Formal relationships between MCDA and production-based performance measurement*

A weak linkage between production-based and multicriteria evaluation also exists in methodological papers that compare DEA (as a method based on production theory) with MCDA in general.[19] This literature, mostly published in the 1990s or before, considers DEA either as a MCDA method itself or as a topic of performance measurement of past actions which is independent from the topic of facilitating future decisions by MCDM. Although MCDM and DEA originated at much the same time in the 1970s, they developed largely independently in the first decades (cf. Belton & Stewart (2002), p. 298), probably not only because of their different purposes but also due to the fact that DEA renders a priori weighting of multiple criteria by the decision maker unnecessary. Belton (1992) and Doyle & Green (1993) were among the first who described the formal relationships between DEA and MCDA as independent topics. Belton & Vickers (1993), Stewart (1996), and Belton & Stewart (1999) put forward the suggestion that an MCDA interpretation of DEA can facilitate understanding. Joro, Korhonen & Wallenius (1998) detailed the mathematical relationship and close similarities between DEA and MOLP using the reference point approach (cf. Wallenius et al. (2008), p. 1343).

By interpreting distance functions used in DEA as measures of the preferences regarding the inputs and outputs of the production activities of distinct DMUs, the framework of Kleine (2004) utilises scalarising functions from MCDA in order to develop classification schemes and generalisations of more or less standard DEA models (whereas the generalisation of DEA models by Yu, Wei & Brockett (1996) is predominantly mathematically motivated). Liu, Sharp & Wu (2006) identify three key building blocks in each DEA model, namely preference order, PPS, and performance measure. Their framework also allows to classify and extend most well-known DEA model types.

The literature of Category 2b is important as it reveals similarities between DEA and MCDA that pave the way for a stronger integration (Category 3). Unfortunately, this kind of study is rarely noticed in most of the recently published application-driven DEA literature. Therefore, analysing the principal connections between production-based efficiency measurement and MCDA seems to be a topic of little interest in the current literature.

#### 4.2.3. Strong combination of MCDA and production theory (Category 3)

Analogously to main Category 2, the third one also assesses a finite set $\mathcal{E}$ of observed production activities by analysing an infinite 'data envelopment' $\mathcal{A}$. Going beyond pure 'technical efficiency' analysis implies the introduction of either (more) information about preferences or arbitrariness[20] (Bouyssou, 1999). This has to be strictly distinguished from cases where the PPS $\mathcal{P}$ or data envelopment $\mathcal{A}$ is not definitely determined a priori so that it can be further enlarged, e.g. by additional technological information about substitution possibilities between individual inputs or outputs in their transformation process, called "production trade-offs" (Podinovski, 2015).

The third category encompasses methodologies that combine preference-based and production-based performance evaluation in a closer, more compatible fashion, be it implicitly or explicitly. Again, two subcategories can be distinguished, depending either on (a) whether they accept the usual premises of DEA without any further discussion and incorporate more or less value information about the trade-off between the inputs and outputs or (b) whether they first aggregate several input and output types – for which production theoretical relations are explicitly stated – in a consistent (perhaps hierarchical) preference-based manner before applying DEA or other appropriate evaluation methods.

Category 3a: *Incorporating preference information or MCDA methods into common DEA*

There is a comprehensive set of OR/MS literature which starts from DEA as a production-based research strand by (implicitly) assuming some information on the trade-off between the inputs and outputs of the compared activities – even if this literature often refrains from explicitly addressing the terms of MCDA topic (B)

[18] The wide range of applications and methodical combinations of multicriteria performance evaluation with DEA is well demonstrated by the article of Eilat, Golany & Shtub (2008) which we found in Step 3 of our search when combining multicriteria performance evaluation (A, B) with the accounting topic (D). For the purpose of R&D project evaluation, it connects DEA with the Balanced Scorecard, an instrument usually serving as a strategic management tool for the hierarchical organization of a company's multiple objectives, however, without taking a closer look at production alternatives.

[19] Our search predominantly revealed papers comparing DEA with MCDA in general, but also one comparing DEA ranking with specific MCDM tools (Sarkis, 2000).

[20] Certain approaches to cope with undesirable outputs in DEA are examples of such arbitrariness. Wojcik, Dyckhoff & Gutgesell (2017) find it striking that all radial models with data transformation for the input or output quantities of bads must assume variable returns to scale in order to avoid distortions of the efficiency evaluation of the DMUs. Typically, such data transformations multiply the quantities of bads by −1 and add the same large enough constant to all those quantities in order to achieve positive numbers. Then, the assumption of variable returns to scale is purely mathematically motivated and does not need to conform with the actual properties of the real production possibilities underlying the transformation processes of the considered DMUs. In empirical sciences, data transformations have to be justified by arguments of measurement theory with respect to the application area at hand (Roberts, 1979). To neglect this important requirement is a serious deficit of all DEA models applying the core idea of data transformation to measure the performance of DMUs without the necessary reflections, particularly in cases with the input or output of bads.

used in Section 4.1. These articles seek to solve the problem that the flexible weighting scheme of DEA calculations may lead to implausible results. Pioneering articles, implementing absolute or relative weight restrictions into DEA models, are published by Dyson & Thanassoulis (1988) and Wong & Beasley (1990). Several further approaches were developed mainly in the 1990s with the aim of a more balanced[21] and preference-oriented performance evaluation. They are predominantly concerned with weight restrictions, target setting or other kinds of incorporating value judgements into DEA. Besides an early review by Allen et al. (1997), this topic is explicitly dealt with in dedicated chapters of many books on DEA (e.g. Cooper, Seiford & Tone (2007) and Joro & Korhonen (2015)).

Typically, these articles are primarily written from a technical point of view, e.g. when solving problems like infeasibility or when discussing the formal compatibility of weight restrictions and additional virtual alternatives (Thanassoulis & Allen, 1998). Methods like the common weights or the cross-efficiency approach limit the individual weights by taking into account the weights of other DMUs (Doyle & Green, 1994). Even if these approaches objectify the evaluation process and “aim at increasing discrimination and producing rankings of efficient units”, they are purely data-driven and “do not incorporate preference information” (Joro & Korhonen (2015), p. 67). This also holds for the so-called ‘multiple criteria data envelopment analysis’ (e.g. Li & Reeves (1999), Hatami-Marbini & Toloo (2017)), which uses distances between DMUs to derive additional efficiency measures that are subsequently optimised in a multi-objective approach.

Other approaches lead to a more direct inclusion of the decision maker’s preferences, often by linking DEA and MCDA in an interactive and sometimes visually supported way (Belton & Vickers (1993)). An example is the target setting approach, pioneered by Golany (1988), which uses MOLP in order to determine the distance of a DMU to an ideal point derived from a decision maker’s preferences (extended to undesirable outputs by Ebrahimnejad, Tavana, & Mansourzadeh (2015)). Related approaches also look at the efficient frontier for an optimal solution (Wong, Luque, & Yang, 2009).

These methods are closely connected to “value efficiency analysis” (VEA), which is one of the most elaborated approaches of this category, introduced by Halme et al. (1999) with the intention of combining DEA and MCDA more strongly. They developed a theory and procedures for complementing efficiency measurement with preference information regarding the desirable structure of inputs and outputs. Based on the assumption of a pseudo-concave preference function, the ‘most preferred solution’ (as vector of input and output quantities) on the efficient frontier is explicitly located. The VEA approach is comprehensively explained in the book by Joro & Korhonen (2015). In the past two decades, it has been further developed and applied to several specific topics of performance evaluation (see Gerami (2019) and the papers of Halme, Joro, Korhonen and co-authors in Table A1).

Category 3b: *Partial aggregation of inputs and outputs by impacts or preferences*

In order to deal with (undesirable) bads as inputs or outputs in the context of ecological efficiency, Dyckhoff & Allen (2001) introduced an approach to multicriteria efficiency measurement based on the decision-theoretical generalisation of traditional production theories by Dyckhoff (1992). In this way, a generalisation of DEA is achieved by using multiple value functions for the relevant objectives that are defined on the quantities of relevant types of inputs and outputs (see Dyckhoff & Souren (2020) for a comprehensive presentation). This approach and its further development (partly in German business economics and not yet documented in the Web of Science) are rarely noticed in the literature until now. Although the article by Dyckhoff & Allen (2001) itself is the most frequently cited of all sources combining topics (A) and (B) with (C) or (F), these citations are nearly always limited to the topic of ecological efficiency, which, however, represents only a specific aspect of the general approach presented in the article. This is why Section 5 outlines insights and main findings of the literature dedicated to this category.

[21] Balance can not only be integrated into the calculation of the DMU’s efficiency score; it can also be measured as a separate index of specialisation complementing the efficiency score (Dyckhoff, Mbock, & Gutgesell, 2015).

### 4.3. Discussion of the literature (not) found

Taking into account ecological or environmental objectives by incorporating bads as outputs (and sometimes also as inputs) has been one of the main strands of development of efficiency measurement in OR/MS literature in the past two decades (Liu et al. (2013a) & (2013b), Lampe & Hilgers (2015)). Special surveys review the knowledge developed so far (Zhou, Ang & Poh (2008), Song et al. (2012), Wojcik, Dyckhoff & Gutgesell (2017)). Halkos & Petrou (2019) differentiate four possible options for treating undesirable outputs. Each of them “has its benefits and drawbacks which each researcher should take into account at every stage of their research and assess which method is more appropriate to be used” (p. 102). They conclude their review with the assertion that incorporating undesirable outputs has proven to be quite a challenge for researchers working on DEA. The main reason for this may be that the literature on efficiency measurement with bads as inputs or outputs focuses on the production-theoretic foundation, as a rule, and widely ignores the multicriteria nature of this challenge. Although preferences are (implicitly) apparent, neither preferences nor values nor multiple objectives form a central topic in this strand of literature.

Approaches bridging the gap between the multicriteria and the production perspective that are explicitly based on a theoretical reasoning of social or natural sciences (not only of mathematics) are scarce. Notably, this holds for literature of Categories 1, 2a and 3a which is often very application-driven. In their review of early sources of Category 3a, Allen et al. (1997) already concluded that the wide variety of different approaches can be explained by the situation which emerges in real-world applications where some preference information is needed that has dictated the way it has been incorporated (cf. Joro & Korhonen (2015), Ch. 6). However, this does not automatically mean that all those methods are not suitable for any further extension and theoretical foundation.

Contrary to the apparent lack of *explicit* theoretical reasoning, the general framework for production-based multicriteria performance evaluation presented in Section 2 mandatorily presumes a sound foundation on both production and decision theory, in fact on their being integrated by some kind of multicriteria production theory (MCPT), e.g. the one determined by the five fundamental assumptions of Section 2.4. The Category 3b approach of Dyckhoff & Allen (2001) has been developed for the basic case of MCPT where impacts are not considered explicitly and where no assumption regarding any overall preferences or utilities is made (other than the consistency assumption A5). In contrast, the value efficiency approach of Category 3a explicitly uses decision-theoretical concepts and assumes the existence of an overall preference function, but only implicitly refers to the production-theoretical foundation of DEA, on which it is built and which it extends (Joro & Korhonen, 2015).

Analogously, the latter seems to hold true for most other sources found by our search. They extend or modify DEA without any change or discussion of its essential foundation by the envelopment of input and output data based on the PPS concept. The inputs and outputs which are minimised or maximised are not called into question, instead it is assumed that they are predetermined by DEA. As a rule, some preference information, implying a certain extent of acceptable trade-offs between them, is merely added in a variety of different ways in order to obtain more meaningful conclusions.

However, the production-theoretical foundation of DEA is flawed if it is derived and founded solely in line with the pioneering approach of Charnes, Cooper & Rhodes (1978). To be specific, it is absolutely necessary not only to explicitly assume technological information about the underlying processes that transform the inputs into the outputs, but also to justify it empirically in real-world applications. Why this fact is not already implied by the original DEA approach itself will be explained in Section 5.1.

Indeed, it would be a category mistake of a performance evaluation not to clearly distinguish between (objective) technological information about the process of transforming inputs into outputs on the one hand and (subjective) preference information of people valuing these inputs and outputs (perhaps via markets) on the other hand. In the specific context of weight restrictions in DEA (cf. Category 3a in Section 4.2.3), Podinovski (2004) termed this distinction "technology versus value thinking" (p. 1316) and explained that production rates of substitution between inputs or outputs should not be confused with those implied by preferences or markets.[22]

In contrast to many DEA applications in OR/MS literature, *economically oriented literature* usually states technological axioms and further assumptions regarding the PPS. "Let us emphasize that the PPS is the key to link DEA models with the economic theory where two key properties for the PPS are the convexity and free-disposability" (Shen et al. (2019), pp. 341-342). Disposability and other assumptions (e.g. 'Input without Output' and 'No Output without Input'; cf. Färe & Grosskopf (1996), pp. 12-13), may contradict the mass and energy balance principles of physics in case that all material objects of the production process were to be relevant for the question at hand. In particular, the assumption of weak disposability is largely questioned regarding its ability to account for detrimental outputs. Different kinds of disposability assumptions that were developed for environmental performance analysis are reviewed and criticised by Dakpo, Jeanneaux & Latruffe (2016) as well as by Dakpo & Ang (2019). They may have a strong impact not only on the PPS, but even on its efficient frontier. It is important to notice that all sorts of assumptions used in the literature of economics and frontier analysis that change the efficient frontier have to be justified empirically or theoretically in order to obtain valid results of the performance evaluation (Dyckhoff, 2019). Shen et al. (2019) recently generalised the free disposability assumption with respect to arbitrary preferences in such a way that it also generalises the 'extended strong disposability' assumption of Liu, Meng & Zhang (2010) for cases of bads and goods. In fact, these assumptions add only such activities to the production possibilities which do not alter the efficient frontier. If prices do not exist or are not known, economically oriented performance analyses also seem to focus on the relation of inputs and outputs by using various concepts of productivity and efficiency.[23] Nonetheless, this strand of economic literature does not pick out preferences, values, or multiple objectives as a central theme either.[24]

Apparently, there exists no 'Category 4' approach that fully integrates MCDA with production theory for performance evaluation. Nevertheless, the Category 3b approach will provide a systematic foundation for such a complete integration if it is combined with appropriate methods of Category 3a in a future research path.

## 5. Insights and findings of a production-based multicriteria performance analysis

This section outlines and clarifies insights and main findings regarding the data envelopment methodology for multicriteria performance evaluation developed so far by the literature of Category 3b. It puts an entirely new complexion on several open questions and problems discussed in the literature, partly reviewed in the last section.

Without loss of generality, we concentrate on the production process-related value function $\boldsymbol{v}(\boldsymbol{x};\boldsymbol{y})$. A distinction of their two components $\boldsymbol{r}(\boldsymbol{x};\boldsymbol{y})$ and $\boldsymbol{v}=\boldsymbol{w}(\boldsymbol{r})$, with $\boldsymbol{v}=\boldsymbol{w}(\boldsymbol{r}(\boldsymbol{x};\boldsymbol{y}))$ defined in Section 2, is made solely within our verbal reasoning. The overview mainly analyses the influence of different types of value functions $\boldsymbol{v}(\boldsymbol{x};\boldsymbol{y})$, i.e. cost and benefit functions $\boldsymbol{c}(\boldsymbol{x};\boldsymbol{y})$ and $\boldsymbol{b}(\boldsymbol{x};\boldsymbol{y})$, respectively. First, the category mistake of DEA applications, already mentioned before, is demonstrated by a simple numerical example. Then, the relevance of three main propositions that have been derived by Dyckhoff (2018) and (2019) is explained. They are concerned with the convexity, consistency, and linearity of valuations, determined by respective types of non-linear and linear impact or value functions.

### 5.1. Data envelopment of costs and benefits or of inputs and outputs?

Just in line with the quotation of Cook, Tone & Zhu (2014), cited in our introduction, the literature review of Section 4 has shown that DEA is often applied as a multicriteria evaluation method to impacts or values which are then misleadingly called 'inputs' or 'outputs'. For example, the following CCR-models in envelopment form are erroneously used this way by enveloping observed costs and benefits without taking notice of their possible dependence on the actual inputs and outputs:

$$\theta_o^* = \min_{\boldsymbol{\lambda}\geq\boldsymbol{0}} \theta_o \text{ such that } \sum_{j=1}^{n} \lambda_j \boldsymbol{c}_j \leq \theta_o \boldsymbol{c}_o \text{ and } \sum_{j=1}^{n} \lambda_j \boldsymbol{b}_j \geq \boldsymbol{b}_o \tag{6}$$

$$\eta_o^* = \max_{\boldsymbol{\lambda}\geq\boldsymbol{0}} \eta_o \text{ such that } \sum_{j=1}^{n} \lambda_j \boldsymbol{c}_j \leq \boldsymbol{c}_o \text{ and } \sum_{j=1}^{n} \lambda_j b_j \geq \eta_o \boldsymbol{b}_o \tag{7}$$

Model (6) differs from the input-oriented CCR-model (5) of Section 3.3 simply by replacing the symbols for costs and benefits with those for the inputs and outputs: $\boldsymbol{c} \leftrightarrow \boldsymbol{x}$ und $\boldsymbol{b} \leftrightarrow \boldsymbol{y}$. Syntactically, models (5) and (6) are identical. Semantically though, the above LP

[22] Further examples of confusing technology and value thinking are certain ways of treating undesirable factors in DEA (Dyckhoff, 2018). In an overview of "DEA models with undesirable inputs, intermediates, and outputs", Zhou & Liu (2015) define desirable outputs as "what the decision maker hopes to produce as much as possible" (p. 417). This reflects a subjective judgement. In contrast, their notion of 'undesirable inputs' is determined technologically: "the desirability of inputs should be defined according to the intrinsic production mechanism. (…) If the increase of an input will not increase the desirable outputs, then it is classified as undesirable" (p. 417), in the opposite case as desirable. The increase of a limitational (i.e. non-substitutable) production factor, e.g. tyres in car assembly, without increasing other factors at the same time, will neither increase nor decrease the output of cars as main products. Hence, according to the technological definition of Zhou & Liu (2015), limitational factors would have to be classified as inputs which are simultaneously both desirable as well as undesirable. Furthermore, waste incineration plants do not even have any desirable outputs so that their input 'waste' cannot be classified in this technological way.

[23] See e.g. the recent handbook by ten Raa & Greene (2019), in particular its introduction (ten Raa, 2019).

[24] A recent exception with interesting viewpoints was published by Agasisti, Munda & Hippe (2019). Though, Caballero, Romero & Ruiz (2016) wrote in their introduction to a special journal volume on linking economics to MCDM (p. 2): "[T]he acceptability of a theory requires not only its internal coherence, but also a good level of external coherence or correspondence to the factual reality (i.e., a certain degree of empirical corroboration). … In short, MCDM remains somewhat unknown in what can be considered orthodox economics. But it would seem to be totally acceptable that if economic problems would be underpinned by the MCDM optimization theory instead of the classic one, at least some of the lack of external coherence problems pointed out above would be considerably mitigated."

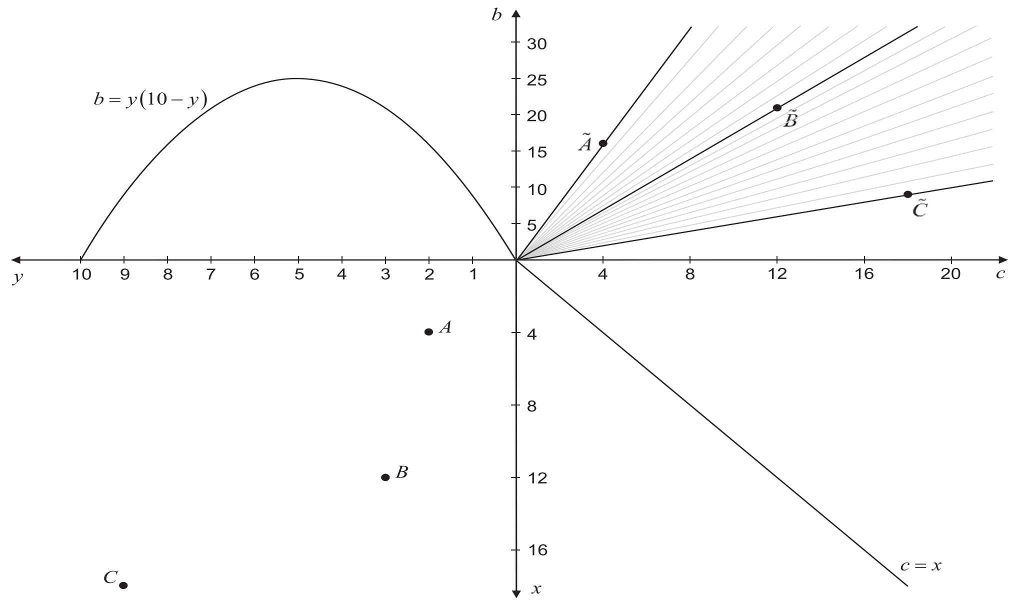


**Fig. 5.** Linear performance preferences regarding costs and benefits (cf. Dyckhoff & Souren (2020), p. 57)

models are fundamentally different from their conventional counterparts (except for the very special case where the costs are indeed measured by the inputs and the benefits by the outputs). This is demonstrated by the following numerical example. It simplifies and modifies the example of Section 2.3 significantly to make it graphically presentable by *Figure 5*.

Let the production activities $\boldsymbol{z} = (x; y)$ of three DMUs $A$, $B$, and $C$ with a single input and a single output be given by $\boldsymbol{z}_A = (4; 2)$, $\boldsymbol{z}_B = (12; 3)$, and $\boldsymbol{z}_C = (18; 9)$, a linear cost function by $c(x; y) = x$ and a quadratic benefit function by $b(x; y) = y(10 - y)$. No further assumptions concerning production possibilities are made. Therefore, only the three activities $A$, $B$, and $C$ exist left below in the $(x,y)$-quadrant. Cost $c$ and benefit $b$ are displayed on the right and top axes of this 4-dimensional coordinate system. The points $\tilde{A}$, $\tilde{B}$, and $\tilde{C}$ as value images of $A$, $B$, and $C$ with $\boldsymbol{v}_A = (4; 16)$, $\boldsymbol{v}_B = (12; 21)$, and $\boldsymbol{v}_C = (18; 9)$ show the associated cost and benefit $\boldsymbol{v} = (c; b)$. Figure 5 also displays the polyhedric cone derived from the left side of the inequalities in LPs (6) and (7) and spanned by the three points in the top-right quadrant as a linear envelopment of their costs and benefits. Nonetheless, $\tilde{A}$, $\tilde{B}$, and $\tilde{C}$ are the only value points that are achievable. Their efficiency scores resulting from both (6) and (7) are $\Theta_A = 1$, $\Theta_B = 7/16$ and $\Theta_C = 1/8$ (with $\eta_o^* = 1/\Theta_o^*$).

The graphs of Figure 5 can be interpreted in classical economic terms based on the well-known *Cournot Theorem* for price fixing by a monopolistic market actor. Analogously to the example in Section 2.3, the DMUs $\rho \in \{A, B, C\}$ may represent distinct business units selling quantities $y_\rho$ of the product on their respective local markets, which are determined by a linear demand function ($y_\rho = 10 - p_\rho$) regarding the individually chosen price $p_\rho$. The upwards-directed axis shows the corresponding revenue $b(y)$. Now, suppose that production takes place with fixed financial costs by exploiting a natural resource as free good in quantity $x$. Nevertheless, the exploitation of the natural resource induces ecological costs $c(x)$ – shown on the axis directed to the right – that cannot be measured in monetary terms. Here, revenues are absolutely bounded by 25 currency units (CU) because of the quadratic benefit function, *irrespective of the actual production possibilities of the DMUs*. In contrast, the benefit-oriented CCR model (7) projects the value points $\tilde{B}$ and $\tilde{C}$ vertically upwards to their (so-called) *target points* (12; 48) and (18; 72) on the ray through the efficient point $\tilde{A}$. The *benchmarks* of 48 CU for the revenue of DMU $B$ and of 72 CU for $C$, calculated by (7), are not achievable at all.

In order to avoid possible misunderstandings and a category mistake one must understand that the polyhedric cone in Figure 5 does not say anything about actual production possibilities. In fact, the linear envelopment of costs and benefits of the DMUs stems from the pioneer contribution of Charnes, Cooper & Rhodes (1978) in which they introduce and justify DEA based on economic and engineering categories. Analogously to their reasoning, you can derive (6) and (7) – as well as the corresponding CCR models in multiplier form – from a syntactically identical optimisation programme[25] which maximises the following quotient $\Theta$ as a top KPI[26] with respect to two sets of weighting factors $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_k)$ and $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_\ell)$ such that each of them allows for a separate linear aggregation of the different costs and the benefits (cf. Dyson et al. (2001), p. 253, regarding this linearity assumption as pitfall in traditional DEA):

$$\Theta := \frac{B}{C} = \frac{\boldsymbol{\mu} \cdot \boldsymbol{b}}{\boldsymbol{\eta} \cdot \boldsymbol{c}} = \frac{\sum_{\beta=1}^{\ell} \mu_\beta b_\beta}{\sum_{\kappa=1}^{k} \eta_\kappa c_\kappa} \tag{8}$$

The top KPI (8) already presumes a linearity of the preferences with respect to the various cost types on the one hand as well as the benefit types on the other hand. It becomes obvious by the fact that any multiplication of either all costs or else all benefits or of both with the same factor leads to the same overall performance ranking of the DMUs if measured by the quotient.

[25] Alternatively to the mathematical transformation of quotient programmes into LPs, well-known from Charnes & Cooper (1962), the empirically motivated derivation by Dyckhoff & Souren (2020), pp. 52–58, uses both degrees of freedom (that are determined by the level of measurement of the two sets of weights on ratio scales) for an adequate normalisation each (cf. Belton & Stewart (2002), pp. 299-300).

[26] The quotient (8) characterises the *value productivity* of the considered activity $\boldsymbol{z} = (\boldsymbol{x}, \boldsymbol{y})$ as a one-dimensional performance indicator $\Theta$ of its multiple values $\boldsymbol{v}(\boldsymbol{z}) = (\boldsymbol{c}, \boldsymbol{b})$ by calculating the extent to which (e.g. economic) benefits arise in relation to (e.g. ecological) costs when inputs are transformed into outputs. In an ecological context, such a quotient of economic value added and ecological damage is called *eco-efficiency* (Kuosmanen & Kortelainen, 2005).

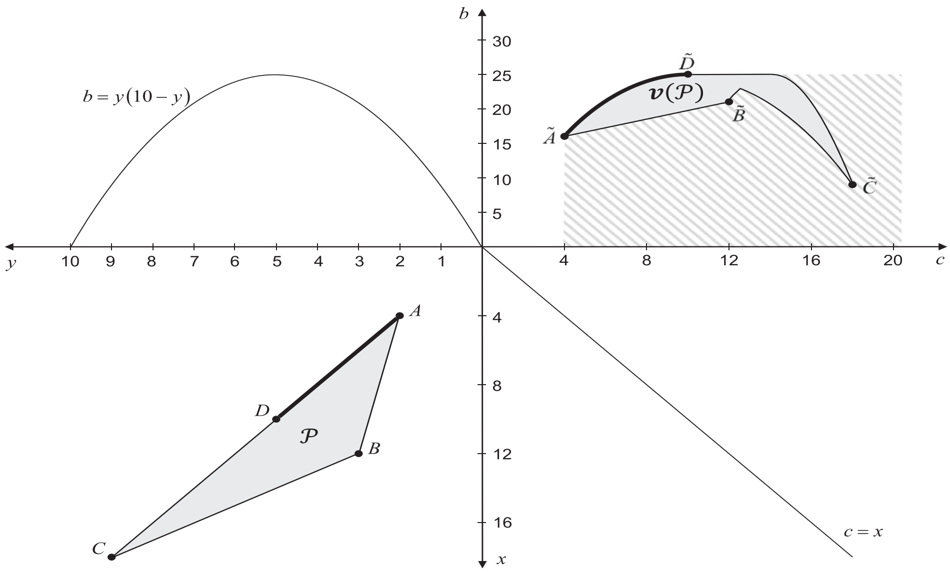


**Fig. 6.** Convex non-linear valuation if values are disposable (cf. Dyckhoff (2019), p. 726)

Indeed, models (6) and (7) permit performance comparisons of all cost/benefit-vectors. Notwithstanding, the example demonstrates that their solutions are hypothetical in general. Regarding the upper-right quadrant of Figure 5, the performance of a point in this space of cost and benefit is uniquely determined by the slope of the corresponding ray through this point. To conclude whether the (target) points on the ray through point $\tilde{A}$ are accessible in reality, additional *explicit* knowledge about the PPS is needed. For instance, *Figure 6* complements Figure 5 by supposing a PPS $\mathcal{P}$ that is determined as the triangle in the below-left quadrant generated by the convex envelopment of the production activities of the three DMUs *A*, *B*, and *C* for a technology with variable returns to scale. As can be seen by its shaded shape, the VPS $\boldsymbol{v}(\mathcal{P})$, representing the value image of the PPS, does not form a convex set in this example because of the non-linearity of the (quadratic) benefit function.

Only those combinations which are on the bold part on the upper left curve of the non-convex shape in Figure 6 are efficient with respect to benefit and cost. This bold curve in the top-right quadrant represents the value image of the bold line between *A* and *D* of the triangle below left. Hence, although the whole line segment between *A* and *C* is traditionally regarded as 'technically efficient' (assuming an input to be minimised and an output to be maximised), only its bold sub-segment spanning from *A* up to the point of maximal benefit – i.e. points *D* and $\tilde{D}$ in Figure 6 with $\boldsymbol{z}_D = (10; 5)$ – is indeed efficient regarding benefit and cost.

### 5.2. Important properties of impact and value functions

The example shown by Figures 5 and 6 demonstrates that, in order to draw any conclusion on whether certain value points are in fact attainable by activities of the DMUs, you generally have to know the actual VPS $\boldsymbol{v}(\mathcal{P})$, i.e. the PPS $\mathcal{P}$ as well as the production process-related value function $\boldsymbol{v}(\boldsymbol{z})$ for $\boldsymbol{z} \in \mathcal{P}$ (or at least for $\boldsymbol{z} \in \mathcal{A} \subset \mathcal{P}$). In contrast to (6), the following cost-oriented optimisation model is an adequate generalisation of the input-oriented CCR model (5) – as will be explained in this subsection:

$$\theta_o^* = \min \theta_o \text{ such that } \boldsymbol{c}(\boldsymbol{z}) \le \theta_o \boldsymbol{c}(\boldsymbol{z}_o) \text{ and } \boldsymbol{b}(\boldsymbol{z}) \ge \boldsymbol{b}(\boldsymbol{z}_o) \text{ for } \boldsymbol{z} \in \mathcal{A} \subset \mathcal{P} \tag{9}$$

Applied to the numerical example of Section 2.3 with $\mathcal{A} = \{\boldsymbol{z} = \sum_{j=1}^{n} \lambda_j \boldsymbol{z}_j | \boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n) \ge \boldsymbol{0}\}$ as linear envelopment of the four cement plants according to (3), one obtains an optimisation task, here e.g. in case of the first DMU, which is linear except for a strictly concave benefit function:

$$\begin{aligned}
\theta_o^* &= \min \theta_o \text{ such that} \\
c_1 &= y_2 + 8500 y_3 \le 5220 \cdot \theta_o \\
c_2 &= y_3 \le 0.6 \cdot \theta_o \\
b_1 &= 40 y_1 (10 - y_1) - 10 x_1 - 50 x_2 + 20 x_3 \ge 270 \\
b_2 &= x_1 \ge 4 \\
x_1 &= 4\lambda_1 + 4\lambda_2 + 5\lambda_3 + 3\lambda_4 \\
x_2 &= 3\lambda_1 + 5\lambda_2 + 5\lambda_3 + 5\lambda_4 \\
x_3 &= 5\lambda_1 + \lambda_2 + 3\lambda_3 + 3\lambda_4 \\
y_1 &= \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 \\
y_2 &= 120\lambda_1 + 40\lambda_2 + 100\lambda_3 + 100\lambda_4 \\
y_3 &= 0.6\lambda_1 + 0.2\lambda_1 + 0.5\lambda_1 + 0.5\lambda_4 \\
&\lambda_1, \lambda_2, \lambda_3, \lambda_4 \ge 0
\end{aligned}$$

It shows benefit $b_1$ as a non-linear function of the activity levels $\lambda_j$. This differs from model (6) where $\lambda_j$ is an identical weighting factor for all costs and benefits of DMU *j*. Nevertheless, the objective $\theta_o$ of model (9) can be interpreted in the same way as in model (6); it determines the smallest factor to which all individual costs can be reduced proportionately within the set of allowed activities without reducing the benefits of the DMU under consideration. Model (9) generalises the traditional DEA model (5) and demonstrates its differences to the often erroneously used one (6) in cases where the value functions are non-linear such that the duality theory of linear programming is not applicable.

The example further illustrates that the individual costs and benefits of the multi-dimensional functions $\boldsymbol{c}(\boldsymbol{z})$ and $\boldsymbol{b}(\boldsymbol{z})$ are not necessarily separable or disjoint with respect to the different inputs and outputs, neither in general nor in the particular case of model (9). It must therefore be emphasised that it is not advisable to analyse examples such as the one above with model (5) or model (6) instead of (9), even in the special case (of Section 5.2.3) that all cost and benefit functions would be linear.

#### 5.2.1. Convexity of valuations

Convexity of the set of feasible solutions is a property of high importance for the purpose of solving optimisation tasks. If $\boldsymbol{c}(\boldsymbol{z})$

and $\boldsymbol{b}(\boldsymbol{z})$ are (multiple) convex cost and concave benefit functions defined on a convex PPS $\mathcal{P}$, the efficient frontier of VPS $\boldsymbol{v}(\mathcal{P})$ is convex regarding the costs and concave regarding the benefits, too (Dyckhoff (2019), p. 725).

As already the simple example of Figure 6 shows, the VPS itself is not necessarily convex, though. Thus, a convex combination of realisable cost-benefit vectors might not be realisable itself. For example, all convex combinations of the images $\tilde{A}$ and $\tilde{C}$ of the 'technically efficient' points $A$ and $C$ are not attainable by any activity in the PPS. Nonetheless, this problem can be avoided if the following property, called *value disposability*, can be supposed. It allows an extension $\widehat{\boldsymbol{v}(\mathcal{P})}$ of the VPS such that a convex set is formed (Dyckhoff (2019), p. 727), although the PPS itself did not change:[27]

$$\widehat{\boldsymbol{v}(\mathcal{P})} = \left\{ (\boldsymbol{c};\boldsymbol{b}) \in \mathbb{R}_+^{k+\ell} \,\middle|\, \boldsymbol{c} \geq \boldsymbol{c}(\boldsymbol{z}), \boldsymbol{b} \leq \boldsymbol{b}(\boldsymbol{z}),\ \boldsymbol{z} = (\boldsymbol{x};\boldsymbol{y}) \in \mathcal{P} \right\} \quad (10)$$

Figure 6 shows the amendment as hatched area on the right and below the VPS. All benefits attainable by a possible production can be reduced and all costs induced by production activities can be augmented, i.e. all value points dominated by the original VPS are realisable, too.

Value disposability presupposes that, in reality, there exist certain additional activities which are feasible for the considered DMUs, although not modelled explicitly in input/output-terms. Hence, they do not belong to the PPS, i.e. they are not represented by the production technology and the actual restrictions. Instead, they are elements of that part of the production environment which a DMU can yet influence. In the example of Section 5.1, this might be a monopolist donating money, received as revenue for the product, to a charity organisation without further reward. Or the augmentation of ecological cost can result from a larger damage when exploiting nature by extracting more of the resource than what is used as input by the DMU. Value disposability does neither change the efficient frontier nor influence the performance score.

### *5.2.2. Monotonicity and consistency of valuations*

In our above numerical examples, the profit of the DMUs is represented by a quadratic benefit function that is indeed concave, but not monotonous. Non-monotonic valuations might lead to a profit maximum that is technically inefficient. Markets for non-storable goods or all-units quantity discounts may show such real-life situations.

This is, however, not the case if the valuation is *(preferentially) consistent*, that is, if increasing benefits and decreasing costs of a lower level imply increasing benefits on the higher level, and the opposite holds true for the costs. Let $\boldsymbol{v}^1(\boldsymbol{z})$ and $\boldsymbol{v}^2(\boldsymbol{z})$ with $\boldsymbol{v}^2(\boldsymbol{z}) = \boldsymbol{f}(\boldsymbol{v}^1(\boldsymbol{z}))$ be two multiple value functions, whereby $\boldsymbol{f}(\boldsymbol{v})$ is a strictly monotonic function mapping the first-level costs and benefits determined by $\boldsymbol{v}^1(\boldsymbol{z})$ consistently onto the second-level costs and benefits determined by $\boldsymbol{v}^2(\boldsymbol{z})$. Then, the following holds true (Dyckhoff (2018), p. 866): If activity $\boldsymbol{z}_{\mathrm{A}} \in \mathcal{P}$ of DMU A dominates activity $\boldsymbol{z}_{\mathrm{B}} \in \mathcal{P}$ of DMU B with respect to the first value level, A dominates B with respect to the second value level, too. Thus, a monotonic valuation implies that an activity is already efficient on all lower levels if it is efficient regarding a higher valuation level. Moreover, consistent, monotonically nested, multi-stage value or impact functions imply non-improving performance ratings of each DMU in each valuation step (cf. Dyckhoff (2018), p. 872–873, for a numerical example).

[27] This feature distinguishes value disposability from strong disposability generalised by Shen et al. (2019). Nevertheless, an analogous value-concerned generalisation of the traditional axioms 'Input without Output' and 'No Output without Input' by 'Costs without Benefits' and 'No Benefits without Costs' (Dyckhoff, 1992) may have implications for the PPS.

Therefore, it cannot be rational to produce inefficiently if the performance analysis applies preferentially consistent monotonic valuations. This throws a better light on the discussion of the rationality of inefficient production (initiated by Bogetoft & Hougaard (2003); cf. Dyckhoff (2018), p. 877). For example, if input slacks are allocated a positive value, the term 'technical efficiency' loses its importance for a performance analysis that makes sense. Hence, the above consistency proposition questions 'rationalising inefficiency', a topic identified by Avkiran & Parker (2010) a decade ago as one of four directions for future DEA studies.

### *5.2.3. Linearity of valuations*

In the linear case, the duality theory of linear programming – which is essential for DEA – is applicable. Linear functions are simultaneously convex and concave as well as monotonous so that the assertions of the latter two subsections are true for this specific type of impact and value functions. Thus, if a multicriteria DEA model, e.g. that of type (9), is consistently aggregated by linear value functions into a DEA model of the same type on a higher hierarchy level, then the efficiency scores of the DMUs cannot improve (Dyckhoff (2018), p. 871). Most notably, an inefficient activity remains inefficient.

If $\boldsymbol{v}(\boldsymbol{z})$ are multiple linear value functions defined on a convex PPS $\mathcal{P}$, *i.e.* $\boldsymbol{v}(\lambda_1\boldsymbol{z}_1+\lambda_2\boldsymbol{z}_2) = \lambda_1\boldsymbol{v}(\boldsymbol{z}_1)+\lambda_2\boldsymbol{v}(\boldsymbol{z}_2)$ for all $\boldsymbol{z}_j \in \mathcal{P},\ \lambda_j \geq 0,$ $j \in \{1, 2\}$, then $\boldsymbol{v}(\mathcal{P})$ is a convex set, too. Moreover, if $\boldsymbol{v}_j := \boldsymbol{v}(\boldsymbol{z}_j) \in \mathbb{R}^{k+\ell}$ for $\boldsymbol{z}_j = (\boldsymbol{x}_j;\boldsymbol{y}_j) \in \mathbb{R}^{m+s}$ and

$$\mathcal{P} = \left\{ \boldsymbol{z} = \sum_{j=1}^{n} \lambda_j \boldsymbol{z}_j | \boldsymbol{\lambda} \in S \right\}$$

with the activity levels set $S \subset \mathbb{R}_+^n$, then the VPS has the same property in value space (Dyckhoff (2018), p. 868):

$$\boldsymbol{v}(\mathcal{P}) = \{\boldsymbol{v} = \sum_{j=1}^{n} \lambda_j \boldsymbol{v}_j | \boldsymbol{\lambda} \in S\}$$

That is, the image of a convex (or linear) envelopment of activities equals the convex (linear) envelopment of the value image points of these activities. Then, neither the explicit knowledge of the relevant inputs and outputs nor that of the respective linear value functions are necessary to determine the efficiency scores of the DMUs. The linear case $S = \mathbb{R}_+^n$, implying a PPS with *constant returns to scale*, as well as the convex case $S = \{\boldsymbol{\lambda} \in \mathbb{R}_+^n | \sum_{j=1}^{n} \lambda_j = 1\}$ for *variable returns to scale* play eminent roles in DEA. Models (6) and (9) are equivalent in such circumstances.

The above proposition is of fundamental importance for applications of DEA. If its premises hold true in an actual instance, the good news is that it will suffice to know solely the relevant costs and benefits of the DMUs as data observed. This provides a factual justification of the usual DEA LP models when applied directly to the costs and benefits (as so-called 'inputs' and 'outputs'), *provided that the above premises are fulfilled.*[28] As a consequence, the

[28] It is not always made crystal clear in the OR/MS literature on DEA that the target points for inefficient DMUs may otherwise deliver unrealistic benchmarks. On the contrary, sometimes the opposite is suggested, so e.g. by Cooper, Seiford & Zhu (2011), p. 1-2: "Because it requires very few assumptions, DEA has also opened up possibilities for use in cases that have been resistant to other approaches because of the complex (often unknown) nature of the relations between the multiple inputs and multiple outputs involved in DMUs." Such possibly misleading statements can already be found in the pioneering article of Charnes, Cooper & Rhodes (1978), e.g. on page 434: "Unlike other types of production functions, this one derives from (and is therefore directly applicable to) empirical observations." Comparing goal programming and DEA, Cooper (2005) states on page 6: "These evaluations are obtained directly from the data without requiring explicitly formulated assumptions such as linearity, non-linearity, etc. ..." However, DEA presupposes at least a convex (or concave, respectively), piecewise linear ('best practice') approximation of the production frontier, and moreover linear homogeneity in case of the CCR models.

framework of Section 2 leads to conclusions that do not only generalise the traditional methodology of efficiency measurement, but also disclose DEA's fundamental conception and presuppositions and help to avoid the category mistake.

This is of particular importance for the very special instance of linear value functions applied in environmental performance analyses. Instead of ecological impacts as values, the quantities of inputs and outputs themselves are used as proxies which are easy to measure (Dyckhoff & Allen, 2001). The inputs and outputs of activities $\boldsymbol{z} = (\boldsymbol{z}^G; \boldsymbol{z}^B)$ are separated into four categories, those of goods $\boldsymbol{z}^G = (\boldsymbol{x}^G; \boldsymbol{y}^G)$ and those of bads $\boldsymbol{z}^B = (\boldsymbol{x}^B; \boldsymbol{y}^B)$. The corresponding standard preference assumption of the underlying *production theory with goods and bads* reads: Each input of a good and each output of a bad uniquely defines one corresponding type of cost: $\boldsymbol{c}(\boldsymbol{z}) = (\boldsymbol{x}^G; \boldsymbol{y}^B)$, and vice versa, each output of a good as well as each input of a bad defines one type of benefit: $\boldsymbol{b}(\boldsymbol{z}) = (\boldsymbol{x}^B; \boldsymbol{y}^G)$.

Inserting these simple value functions into the cost-oriented multicriteria model (9) for a linear PPS $\mathcal{P}$, spanned by the observed activities of the DMUs, leads to the following radial DEA model (Wojcik, Dyckhoff, & Gutgesell, 2017):

$$\theta_o^* = \min_{\boldsymbol{\lambda}\geq\boldsymbol{0}} \theta_o \text{ such that } \sum_{j=1}^{n} \lambda_j \boldsymbol{x}_j^G \leq \theta_o \boldsymbol{x}_o^G \text{ and } \sum_{j=1}^{n} \lambda_j \boldsymbol{y}_j^B \leq \theta_o \boldsymbol{y}_o^B$$
$$\text{and } \sum_{j=1}^{n} \lambda_j \boldsymbol{x}_j^B \geq \boldsymbol{x}_o^B \text{ and } \sum_{j=1}^{n} \lambda_j \boldsymbol{y}_j^G \geq \boldsymbol{y}_o^G \tag{11}$$

In model (11), bad output is mathematically described in the same way as good input, and bad input like good output. With respect to MCPT, however, this identity is only of a syntactic and not of a semantic nature. These qualities must not be confused! Primary subjects of efficiency analysis are costs and benefits, not inputs and outputs. Therefore: "Considering pollutants as inputs is not a correct way of modelling pollution-generating technologies" (Dakpo, Jeanneaux & Latruffe (2016), p. 357). Emitted pollutants are undeniably outputs; they are undesirable because of their impacts ('external effects') that induce social costs (also known as 'external costs' in economics). Thus, the integration of MCDA with production theory also helps to understand and to cope with the challenge to incorporate (undesirable) bads as inputs or outputs into DEA as well as into economics (discussed in Section 4.3).

## 6. Conclusions, challenges, and future research paths

Three decades ago, Belton (1992) asserted with respect to the relationship between DEA and MCDA that "the two approaches can be integrated to provide a more effective and easier to understand approach to performance measurement" (p. 71). Our framework and review confirm this statement for a more general scope of validity.

In fact, only relatively few sources try to bridge the gap between the multicriteria and the production perspective of performance evaluation until today. One can distinguish different degrees of strength of the link between both perspectives, starting with pure formal comparisons of both isolated perspectives (no substantial link) and then successively strengthening the link substantially up to a theoretically well-founded complete integration, at last. Although the set of literature, which we found, categorised, and discussed in Section 4, shows various degrees of strength, there is not a single concept or approach amongst it that fully integrates both perspectives. Nevertheless, some promising sources and starting points for a stronger substantial integration exist, thus indicating future research paths by enhancing and improving those known from the categories discussed in Section 4.

First of all, we propose to generally apply approaches of Category 3a – like value efficiency of Halme et al. (1999) or other theoretically founded MCDA methods – to the (possibly stepwise aggregated) impacts and values defined by the framework of Section 2 in combination with the Category 3b approach, outlined in Section 5, instead of applying them merely to the inputs and outputs like in the very particular case of common DEA. In our view, Section 2 presents a framework and theoretical fundament for a full integration of MCDA and production theory for the purpose of performance evaluation but has to be complemented and filled in with specific ideas, concepts, and approaches both of MCDA and production theory.

### 6.1. Theoretical and methodological research paths

Regarding DEA research, Avkiran & Parker (2010) ascertained "a sharp decline in the number of theoretical papers as of 2004, thus suggesting a maturing of the methodology" (p. 2). A mature phase is reached when "key underlying assumptions are no longer challenged" (p. 1). In order to generally lay the foundation for more influential work, they suggest a more dramatic and difficult form of investigation than a mere gap identification. This would be "to successfully challenge a foundation theory, or assumptions thereof, or take an undeveloped original idea and present it in a format that can be easily generalized and applied by others" (p. 1).

In our view, such a challenge for traditional production-based methods of performance evaluation, like DEA or Stochastic Frontier Analysis (SFA), requires – as a matter of principle – a sharp distinction between *three different categories of notions* affected by the production process to be evaluated: (1) the inputs and outputs going into or emerging from the process, mostly determined technologically and easy to observe; (2) the results (consequences, effects, impacts, or outcomes) for the human, social, economic, or natural environments of the production system, mainly influenced by the observed inputs and outputs in an objectively or intersubjectively measurable manner; and (3) the costs and benefits as – often subjective, non-financial, incommensurable – values destroyed or created by the process via its results, evaluated with respect to the preferences of a certain person or authority. Such a strict distinction between important notions facilitates the use of knowledge of different scientific disciplines as well as of practical experience of the application areas concerned – and hopefully contributes to avoid a category mistake.

For example, to help management "open the black box of production", Avkiran & Parker (2010) recommended *network DEA* as one of four directions of research for the past decade (p. 4). Although this path has indeed been followed intensely in the recent OR/MS and economic literature (Kao, 2017), it is somewhat surprising that this important topic apparently did not find much interest in the production and operations management literature. The theoretical and technological knowledge of business economics and production engineering with respect to multi-stage and network processes within these applied disciplines can be helpful to analyse realistic models and propose valid assertions. In particular, it may be necessary to distinguish further between those inputs and outputs that go into or emerge from a specific transformation process as part of the considered production system (*process input or output*) and those entering or leaving the whole production system through its boundaries (*system input or output*). This is of the utmost importance if the considered objects can be stored in an inventory, especially in cases of dynamic performance analyses.

In addition to the consideration of specific production relationships, production-based multicriteria performance analyses must also appropriately capture and include the relevant purposes and objectives. Performance evaluation, as defined in Section 2, assesses the results of purposively rational human actions regarding their effectivity in achieving the intended individual ends as well as regarding their efficiency in balancing these ends with respect to the relevant means employed and the unintended secondary results as far as they are desirable or undesirable. These ends, means, and secondary results represent the values which are decisive for the evaluation. Hence, methods of performance evaluation, in particular those of accounting, DEA, and LCA, can *never be totally value free*.

At most, they can try to use as much objectively or intersubjectively acceptable premises and determinable data as possible. Such data may be prices on markets in financial and management accounting as well as laws of nature or legal and social norms in LCA. They allow to derive acceptable valuations for the trade-off between the ends, means, and secondary results when measuring performance. Often, however, such data do not exist or are not available. Then, the ends, means, and secondary results represent multiple objectives of the decision maker, evaluator, or some other authority the attributes of which are not easily comparable. Therefore, in principle, *performance evaluation constitutes a kind of multicriteria analysis* for which concepts, methods, and decision support systems developed for MCDA in OR/MS can be useful (cf. Cinelli et al. (2020)), especially if they are applied in LCA and moreover integrated with domain-specific production knowledge.

Although there is a growing literature which combines MCDA methods with LCA (cf. Section 4.2), it does, as a rule, not integrate multicriteria decision concepts from OR/MS systematically into the conceptual framework of LCA used in practice (ISO 14040) – which, in turn, largely corresponds to the generic structure of performance evaluation developed in Section 2.2 (Figure 2). Therefore, enhancing the few existing approaches, which combine DEA and LCA in a conceptual, structured manner (Lozano et al. (2009), Iribarren et al. (2010), Vazquez-Rowe et al. (2010)), by further insights and methods of MCDA may be a fruitful future research path. A main research topic should be to trace back the plethora of indicators in applications – like e.g. sustainability supply chain assessment (Ahi & Searcy (2015), Quori, Mujkic & Kraslawski (2018)) – to the most important impacts of the inputs and outputs in order to support the aggregation and valuation process of life cycle impact analysis (LCIA; cf. Hauschild & Huijbregts (2015)) in a way that is decision-theoretically sound.

### *6.2. Empirical and application-oriented research paths*

Management accounting is the process of measuring, analysing, and reporting not only financial, but also non-financial information that helps managers make decisions to fulfil the goals of an organisation (Datar & Rajan, 2018). In view of the ecological and social goals of many organisations nowadays, production-based multicriteria performance evaluation methodologies should be a main topic there, too. A look at advanced textbooks and journals on management accounting reveals, however, that such methods, in particular DEA and LCA, are widely ignored in this literature up to now. It is incomprehensible to us, though, why this is the case. The mathematical requirements cannot be an essential barrier since today management accounting research also uses ambitious mathematics. Regarding LCA, it may be the lack of necessary environmental and ecological knowledge. With respect to the broad application areas of DEA, however, specific knowledge – in addition to the pure methodical one – is only necessary regarding the considered application area. We suspect that the actual reason why *DEA has not been successful in management accounting and control* until now is a lack of convincing success stories (e.g. in a journal like *Interfaces*). Already in the early years of DEA, Nunamaker (1985) has pointed to an important difficulty of its real-world application regarding manipulations of variable selection and data determination which may influence its acceptance in practice. Therefore, he suggested (p. 57): "DEA's reliability could be improved through implementation of standardized accounting and reporting requirements coupled with an extensive audit function."

A decade ago, Avkiran & Parker (2010) suggested "DEA in Practice" as a direction for future DEA studies, asking (p. 3): "For example, given the popularity of DEA in academic journals, how widespread is its use in existing organizations?" Most 'applications' of DEA published in the scientific literature are mere calculations with real data, but they do not demonstrate the actual usefulness of DEA for the potential user. Therefore, a main future research path regarding DEA as well as other production-based performance evaluation methodologies should be to *confirm more convincingly that its application really matters and may improve reality*.

In view of many potentially fruitful application areas of production-based multicriteria performance evaluation methods, it is furthermore desirable to develop systematic, clear, and convincing guidelines which explicitly address the issue of how to define and select the relevant input and output variables as well as the appropriate properties of the PPS from a production point of view. For example, within their unified process for non-parametric performance measurement projects, called "COOPER framework", Emrouznejad & De Witte (2010) do not explain how to empirically verify the adequacy of assumptions about the properties of the underlying production process. Instead, they propose to use a preference-oriented suggestion from Cook & Zhu (2008) in cases in which it is not clear whether a variable should be classified as an input or an output (p. 1579): "If an increase in the value of the variable results in an increase [decrease] in the efficiency score then … it is an output [input] variable."[29]

In contrast, in order to provide better support during the process of defining and identifying relevant input and output types, Afsharian, Ahn & Neumann (2016) proposed to complement existing DEA guidelines – like the COOPER-framework – by an additional goal-oriented phase which determines performance criteria based on the generalised DEA concept of Dyckhoff & Allen (2001) as well as on a general decision-oriented performance measurement framework. Their approach is in line with the protocol drawn up by Dyson et al. (2001) to choose performance measures that are "strongly related to the objectives of the organisation. This might be achieved by a careful consideration of the consistency of the mission, objectives and performance measures" (p. 248). Taxonomies how to conduct the MCDA process in supporting decisions – like that one recently introduced by Cinelli et al. (2020) – may be fruitful in suggesting ideas and concepts for convincingly evaluating performance in practice. Furthermore, methods for problem structuring (Belton & Stewart, 2010), particularly those for defining and structuring fundamental objectives (Eisenführ, Weber & Langer (2010), Ch. 3), can be useful in determining and selecting relevant input and output types and their impacts on the considered environments of a production system. Yet, the challenge how to verify the empirical validity of assumptions about the production properties is not really met until today. We think, however, that such verifications must necessarily be specific for the addressed application area which means that generic guidelines and frameworks alone will not be sufficient.

Thus, the prerequisites to successfully apply quantitative methods of production-based multicriteria performance evaluation in practice are usually very challenging. Although there is a strong demand for such approaches, which often is pointedly expressed by the phrase: "*You can't manage what you can't measure!*", this assertion has to be contrasted with a second well-known phrase which says: "*What you measure is what you get!*" The latter quote points to the severe neglection of those performance criteria which

[29] See also the more recent statement in a DEA handbook of Cook & Zhu (2014), p. vii: "The performance or efficiency of a DMU is expressed in terms of a set of measures which are classified or coined as DEA inputs and outputs."

are not measurable in valid quantitative terms despite their high relevance. To avoid such a dilemma, any convincing performance evaluation should in fact use all relevant valid information. Hence, performance evaluation is more than mere measurement. Nevertheless, it comprises measurement as far as possible. Well-proven quantitative methods should be applied if the prerequisites are fulfilled, even if for parts of the data only. Their results may constitute a *management dashboard* as core of a performance measurement and management system (Dyckhoff & Souren (2020), p. 104).

## Appendix A. Representative literature

**Table A1**
Representative literature found by the four-step search process

| Literature | Search Step* | Citations (counted on 30.11.2020) Web of Science | Google Scholar |
|---|---|---|---|
| **Category 1** | | | |
| Parkan & Wu (1997) | 3 | 50 | 89 |
| Ho, Xu & Dey (2010) | 2 (R) | 1011 | 2309 |
| Brandenburg et al. (2014) | 2 (A) | 523 | 954 |
| Ng & Chuah (2014) | 3 | 19 | 29 |
| Eskandarpour et al. (2015) | 2 (R) | 280 | 474 |
| Ilgin, Gupta & Battaia (2015) | 2 (R) | 49 | 96 |
| Gumus et al. (2016) | 3 | 14 | 22 |
| Banasik et al. (2018) | 2 (R) | 26 | 55 |
| Moons, Waeyenbergh & Pintelon (2019) | 2 (R) | 23 | 72 |
| Thies et al. (2019) | 2 (R) | 20 | 43 |
| **Category 2a** | | | |
| Liu, Huang, & Yen (2000) | 1 | 21 | 69 |
| Eilat, Golany & Shtub (2008) | 3 | 144 | 388 |
| Quariguasi Frota Neto et al. (2008) | 2 (A) | 213 | 556 |
| Madlener, Antunes & Dias (2009) | 1 | 63 | 140 |
| Quariguasi Frota Neto et al. (2009) | 2 (A) | 140 | 266 |
| San Cristobal (2011) | 4 (F) | 53 | 110 |
| Kuo & Lin (2012) | 2 (A) | 99 | 168 |
| Mahdiloo, Saen & Lee (2015) | 2 (A) | 67 | 101 |
| Soheilirad et al. (2018) | 1 | 26 | 45 |
| Agasisti, Munda & Hippe (2019) | 4 (F) | 5 | 10 |
| **Category 2b** | | | |
| Belton (1992) | 2 (A) | – | 63 |
| Doyle & Green (1993) | 2 (A) | – | 241 |
| Stewart (1996) | 3 | 150 | 333 |
| Yu, Wei & Brockett (1996) | 1 | 81 | 145 |
| Joro, Korhonen & Wallenius (1998) | 2 (A) | 132 | 291 |
| Belton & Stewart (1999) | 2 (A) | – | 68 |
| Sarkis (2000) | 3 | 102 | 230 |
| Belton & Stewart (2002)** | 2 (A) | – | 4726 |
| Kleine (2004) | 1 | 39 | 117 |
| Liu, Sharp & Wu (2006) | 1 | 28 | 46 |
| Wallenius et al. (2008) | 2 (R) | 402 | 850 |
| Cook, Tone & Zhu (2014) | 3 | 296 | 602 |
| Tavana et al. (2018) | 3 | 5 | 14 |
| Ehrgott, Hasannasab & Raith (2019) | 4 (F) | 0 | 2 |
| **Category 3a** | | | |
| Golany (1988) | 1 | 190 | 349 |
| Belton & Vickers (1993) | 3 | 89 | 212 |
| Halme et al. (1999) | 2 (A) | 155 | 346 |
| Li & Reeves (1999) | 3 | 202 | 413 |
| Korhonen, Tainio & Wallenius (2001) | 3 | 113 | 314 |
| Joro, Korhonen & Zionts (2003) | 3 | 16 | 39 |
| Korhonen & Syrjänen (2004) | 2 (A) | 156 | 242 |
| Mavrotas & Trifillis (2006) | 3 | 34 | 69 |
| Chen, Larbani & Chang (2009) | 3 | 23 | 53 |
| Yang et al. (2009) | 3 | 14 | 25 |
| Wong, Luque & Yang (2009) | 4 (B) | 26 | 52 |
| Hosseinzadeh Lotfi et al. (2010) | 3 | 10 | 45 |
| Jahanshahloo et al. (2011) | 1 | 12 | 21 |
| Malekmohammadi, Hosseinzadeh Lotfi & Jaafar (2011) | 4 (F) | 8 | 19 |
| Halme, Korhonen & Eskelinen (2014) | 3 | 8 | 17 |

(*continued on next page*)

**Table A1**
(*continued*)

| Literature | Search Step* | Citations (counted on 30.11.2020) Web of Science | Google Scholar |
|---|---|---|---|
| Dyckhoff, Mbock & Gutgesell (2015) | 4 (F) | 2 | 6 |
| Ebrahimnejad, Tavana & Mansourzadeh (2015) | 1 | 3 | 7 |
| Jain et al. (2015) | 3 | 19 | 30 |
| Joro & Korhonen (2015)** | 4 (F) | – | 47 |
| Hatami-Marbini & Toloo (2017) | 3 | 13 | 25 |
| Rubem, Soares de Mello & Angulo Meza (2017) | 2 (A) | 10 | 19 |
| Gerami (2019) | 1 | 1 | 1 |
| **Category 3b** | | | |
| Dyckhoff & Allen (2001) | 1 | 264 | 492 |
| Afsharian, Ahn & Neumann (2016) | 4 (F) | 4 | 17 |
| Dyckhoff (2018) | 4 (F) | – | 11 |
| Dyckhoff (2019) | 4 (F) | – | 2 |

*: (A) = additional source emerging from references of the reviews, (B) = backward search, (F) = forward search, (R) = found in Web of Science as a review
**: Indicates a book

## References


Afsharian, M., Ahn, H., & Neumann, L. (2016). Generalized DEA: An approach for supporting input/output factor determination in DEA. *Benchmarking: An International Journal, 23*, 1892–1909.

Agasisti, T., Munda, G., & Hippe, R. (2019). Measuring the efficiency of European education systems by combining Data Envelopment Analysis and Multiple-Criteria Evaluation. *Journal of Productivity Analysis, 51*, 105–124.

Ahi, P., & Searcy, C. (2015). An analysis of metrics used to measure performance in green and sustainable supply chains. *Journal of Cleaner Production, 86*, 360–377.

Allen, R., Athanassopoulos, A., Dyson, R., & Thanassoulis, E. (1997). Weights restrictions and value judgements in data envelopment analysis. *Annals of Operations Research, 73*, 13–34.

Avkiran, N. K., & Parker, B. R. (2010). Pushing the DEA research envelope. *Socio-Economic Planning Sciences, 44*, 1–7.

Azapagic, A., & Clift, R. (1999). Life cycle assessment and multiobjective optimisation. *Journal of Cleaner Production, 7*, 135–143.

Banasik, A., Bloemhof-Ruwaard, J. M., Kanellopoulos, A., Claassen, G., & van der Vorst, J. G. (2018). Multi-criteria decision making approaches for green supply chains: A review. *Flexible Services and Manufacturing Journal, 30*, 366–396.

Belton, V. (1992). Integrating data envelopment analysis with multiple criteria decision analysis. In A. Goioechea, L. Duckstein, & S. Zionts (Eds.), *Proceedings IXth International Conference on Mutiple Criteria Decision Making* (pp. 71–79). Berlin: Springer.

Belton, V., & Stewart, T. J. (1999). DEA and MCDA: competing or complementary approaches? In N. Meskens, & M. Roubens (Eds.), *Advances in Decision Analysis* (pp. 87–104). Dordrecht: Kluwer.

Belton, V., & Stewart, T. J. (2002). *Multiple Criteria Decision Making: An Integrated Approach*. Norwell MA (USA): Kluwer.

Belton, V., & Stewart, T. J. (2010). Problem structuring and multiple criteria decision analysis. In M. Ehrgott, J. Figueira, & S. Greco (Eds.), *Trends in Multiple Criteria Decision Analysis* (pp. 209–239). Boston: Springer.

Belton, V., & Vickers, S. P. (1993). Demystifying DEA: A visual interactive approach based on multi criteria analysis. *Journal of the Operational Research Society, 44*, 883–896.

Block, J. H., & Fisch, C. (2020). Eight tips and questions for your bibliographic study in business and management research. *Management Review Quarterly, 70*, 307–312.

Bloemhof-Ruwaard, J. M., Koudijs, H. G., & Vis, J. C. (1995). Environmental impacts of fat blends: A methodological study combining life cycle analysis, multiple criteria decsion making and linear programming. *Environmental & Resource Economics, 6*, 371–387.

Bogetoft, P., & Hougaard, J. (2003). Rational inefficiencies. *Journal of Productivity Analysis, 20*, 243–271.

Bouyssou, D. (1999). Using DEA as a tool for MCDM: some remarks. *Journal of the Operational Research Society, 50*, 974–978.

Brandenburg, M., Govindan, K., Sarkis, J., & Seuring, S. (2014). Quantitative models for sustainable supply chain management: Developments and directions. *European Journal of Operational Research, 233*, 299–312.

Caballero, R., Romero, C., & Ruiz, F. (2016). Multiple criteria decision making and economics: An introduction. *Annals of Operations Research, 245*, 1–5.

Campos-Guzman, V., Socorro Garcia-Cascales, M., Espinosa, N., & Urbina, A (2019). Life cycle analysis with multi-criteria decision making: A review of approaches for the sustainability evaluation of renewable energy technologies. *Renewable & Sustainable Energy Reviews, 104*, 343–366.

Charnes, A., & Cooper, W. (1962). Programming with linear fractionals. *Naval Research Logistics Quarterly, 2*, 181–186.

Charnes, A., Cooper, W., & Rhodes, E. (1978). Measuring efficiency of decision-making units. *European Journal of Operational Research, 2*, 429–444.

Chen, Y. W., Larbani, M., & Chang, Y. P. (2009). Multiobjective data envelopment analysis. *Journal of the Operational Research Society, 60*, 1556–1566.

Cinelli, M., Kadzinski, M., Gonzalez, M., & Slowinski, R. (2020). How to support the application of multiple criteria decision analysis? Let us start with a comprehensive taxonomy. *Omega, 96*(102261), 1–22.

Cook, W. D., & Zhu, J. (2008). Classifying inputs and outputs in data envelopment analysis. *European Journal of Operational Research, 180*, 692–699.

Cook, W. D., & Zhu, J. (2014). *Data Envelopment Analysis: A Handbook on the Modeling of Internal Structures and Networks*. New York: Springer.

Cook, W. D., Tone, K., & Zhu, J. (2014). Data envelopment analysis: Prior to choosing a model. *Omega, 44*, 1–4.

Cooper, W. W. (2005). Origins, uses of, and relations between goal programming and data envelopment analysis. *Journal of Multi-Criteria Decision Analysis, 13*, 3–11.

Cooper, W. W., Seiford, L. M., & Tone, K. (2007). *Data Envelopment Analysis* (2nd ed). New York: Springer.

Cooper, W. W., Seiford, L. M., & Zhu, J. (2011). Data envelopment analysis: history, models, and interpretations. In W. W. Cooper, L. M. Seiford, & J. Zhu (Eds.), *Handbook on Data Envelopment Analysis* (pp. 1–39). New York: Springer.

Dakpo, K. H., & Ang, F. (2019). Modelling environmental adjustments of production technologies: A literature review. In T. ten Raa, & W. H. Greene (Eds.), *The Palgrave Handbook of Economic Performance Analysis* (pp. 601–657). Cham: Palgrave MacMillan.

Dakpo, K. H., Jeanneaux, P., & Latruffe, L. (2016). Modelling pollution-generating technologies in performance benchmarking: Recent developments, limits, and future prospects in the nonparametric framework. *European Journal of Operational Research, 250*, 347–359.

Datar, S., & Rajan, M. (2018). *Horngreen's Cost Accounting: A Managerial Emphasis* (16th ed.). Essex: Pearson Prentice Hall.

Demski, J., & Feltham, G. (1976). *Cost Determination: A Conceptual Approach*. Ames: Iowa State University Press.

Doyle, J., & Green, R. (1993). Data envelopment analysis and multiple criteria decision making. *Omega, 21*, 713–715.

Doyle, J., & Green, R. (1994). Efficiency and cross-efficiency in DEA: Derivations, meanings, and uses. *Journal of the Operational Research Society, 45*, 567–578.

Dyckhoff, H. (1992). *Betriebliche Produktion (Business Production)*. Berlin et al.: Springer.

Dyckhoff, H. (2018). Multi-criteria production theory: Foundation of non-financial and sustainability performance evaluation. *Journal of Business Economics, 88*, 851–882.

Dyckhoff, H. (2019). Multi-criteria production theory: Convexity propositions and reasonable axioms. *Journal of Business Economics, 89*, 719–735.

Dyckhoff, H., & Allen, K. (2001). Measuring ecological efficiency with data envelopment analysis (DEA). *European Journal of Operational Research, 132*, 312–325.

Dyckhoff, H., & Souren, R. (2020). *Performance Evaluation: Foundations and Challenges*. Cham (CH): SpringerNature.

Dyckhoff, H., Mbock, E., & Gutgesell, S. (2015). Distance-based measures of specialization and balance in multi-criteria: A DEA-integrated method. *Journal of Multi-Criteria Decision Analysis, 22*, 197–212.

Dyckhoff, H., Quandel, A., & Waletzke, K. (2015). Rationality of eco-efficiency methods: Is the BASF analysis dependent on irrelevant alternatives? *International Journal of Life Cycle Assessment, 20*, 1557–1567.

Dyson, R. G., & Thanassoulis, E. (1988). Reducing weight flexibility in data envelopment analysis. *Journal of the Operational Research Society, 39*, 563–576.

Dyson, R. G., Allen, R., Camanho, A. S., Podinovski, V. V., Sarrico, C. S., & Shale, E. A. (2001). Pitfalls and protocols in DEA. *European Journal of Operational Research, 132*, 245–259.

Ebrahimnejad, A., Tavana, M., & Mansourzadeh, S. M. (2015). An interactive MOLP method for solving output-oriented DEA problems with undesirable factors. *Journal of Industrial and Management Optimization, 11*, 1089–1110.

Ehrgott, M., Hasannasab, M., & Raith, A. (2019). A multiobjective optimization approach to compute the efficient frontier in data envelopment analysis. *Journal of Multi-Criteria Decision Analysis, 26*, 187–198.

Eilat, H., Golany, B., & Shtub, A. (2008). R&D project evaluation: An integrated DEA and balanced sorecard approach. *Omega, 36*, 895–912.
Eis, D., Helm, D., Laußmann, D., & Stark, K. (2010). *Klimawandel und Gesundheit: Ein Sachstandsbericht (Climate Change and Health: State-of-the art)*. Berlin: Robert Koch-Institut.
Eisenführ, F., Weber, M., & Langer, T. (2010). *Rational Decision Making*. Berlin: Springer.
Emrouznejad, A., & De Witte, K. (2010). COOPER-framework: A unified process for non-parametric projects. *European Journal of Operational Research, 207*, 1573–1586.
Eskandarpour, M., Dejax, P., Miemczyk, J., & Peton, O. (2015). Sustainable supply chain network design: An optimization-oriented review. *Omega, 54*, 11–32.
Ewert, R., & Wagenhofer, A. (2014). *Interne Unternehmensrechnung (Management accounting)* (8th ed.). Berlin et al.: Springer.
Färe, R., & Grosskopf, S. (1996). *Intertemporal Production Frontiers: With Dynamic DEA*. Boston: Kluwer.
Farrell, M. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society Series A, 120*, 253–290.
Fisch, C., & Block, J. (2018). Six tips for your (systematic) literature review in business and management research. *Management Review Quarterly, 68*, 103–106.
French, S. (1986). *Decision Theory: An Introduction to the Mathematics of Rationality*. Chichester: Ellis Horwood.
Frisch, R. (1965). *Theory of Production*. Dordrecht: D. Reidel Publ.
Gerami, J. (2019). An interactive procedure to improve estimate of value efficiency in DEA. *Expert Systems with Applications, 137*, 29–45.
Golany, B. (1988). An interactive MOLP procedure for the extension of DEA to effectiveness analysis. *Journal of the Operational Research Society, 39*, 725–734.
Grosse-Sommer, A. P., Grünenwald, T. H., Paczkowski, N. S., van Gelder, R. N., & Saling, P. R. (2020). Applied sustainability in industry: The BASF eco-eEfficiency toolbox. *Journal of Cleaner Production*, (120792), 258.
Guinee, J. B., Heijungs, R., Huppes, G., Zamagni, A., Masoni, P., Buonamici, R., …, & Ryedberg, T. (2011). Life cycle assessment: Past, present, and future. *Environmental Science and Technology, 45*, 90–96.
Gumus, S., Egilmez, G., Kucukvar, M., & Park, Y. S. (2016). Integrating expert weighting and multi-criteria decision making into eco-efficiency analysis: The case of US manufacturing. *Journal of the Operational Research Society, 67*, 616–628.
Halkos, G., & Petrou, K. N. (2019). Treating undesirable outputs in DEA: A critical review. *Economic Analysis and Policy, 62*, 97–104.
Halme, M., Joro, T., Korhonen, P., Salo, S., & Wallenius, J. (1999). A value efficiency approach to incorporating preference information in data envelopment analysis. *Management Science, 45*, 103–115.
Halme, M., Korhonen, P., & Eskelinen, J. (2014). Non-convex value efficiency analysis and its application to bank branch sales evaluation. *Omega, 48*, 10–18.
Hatami-Marbini, A., & Toloo, M. (2017). An extended multiple criteria data envelopment model. *Expert Systems with Applications, 73*, 201–219.
Hauschild, M., & Huijbregts, M. (2015). *Life Cycle Impact Assessment*. New York: Springer.
Ho, W., Xu, X., & Dey, P. K. (2010). Multi-criteria decision making approaches for supplier evaluation and selection: A literature review. *European Journal of Operational Research, 202*, 16–24.
Hosseinzadeh Lotfi, F., Jahanshahloo, G. R., Soltanifar, M., Ebrahimnejad, A., & Mansourzadeh, S. M. (2010). Relationship between MOLP and DEA based on output-oriented CCR dual model. *Expert Systems with Applications, 37*, 4331–4336.
Ilgin, M. A., Gupta, S. M., & Battaia, O. (2015). Use of MCDM techniques in environmental conscious manufacturing and product recovery: State of the art. *Journal of Manufacturing Systems, 37*, 746–758.
Iribarren, D., & Vazquez-Rowe, I. (2013). Is labor a suitable input in LCA + DEA studies? Insights on the combined use of economic, environmental and social parameters. *Social Sciences, 2*, 114–130.
Iribarren, D., Vazquez-Rowe, I., Moreira, M., & Feijoo, G. (2010). Further potentials in the joint implementation of life cycle assessment and data envelopment analysis. *The Science of the Total Environment, 408*, 5265–5272.
Jahanshahloo, G., Zohrehbandian, M., Alinezhad, A., Naghneh, S., Abbasian, H., & Mavi, R. (2011). Finding common weights based on the DM's preference information. *Journal of the Operational Research Society, 62*, 1796–1800.
Jain, V., Kumar, A., Kumar, S., & Chandra, C. (2015). Weight restrictions in Data Envelopment Analysis: A comprehensive Genetic Algorithm based approach for incorporating value judgements. *Expert Systems with Applications, 42*, 1503–1512.
Joro, T., & Korhonen, P. (2015). *Extension of Data Envelopment Analysis with Preference Information*. New York: Springer.
Joro, T., Korhonen, P., & Wallenius, J. (1998). Structural comparison of data envelopment analysis and multiple objective linear programming. *Management Science, 44*, 962–970.
Joro, T., Korhonen, P., & Zionts, S. (2003). An interactive approach to improve estimates of value efficiency in data envelopment analysis. *European Journal of Operational Research, 149*, 688–699.
Kao, C. (2017). *Network Data Envelopment Analysis: Foundations and Extensions*. Cham: Springer.
Kleine, A. (2004). A general model framework for DEA. *Omega, 32*, 17–23.
Koopmans, T. C. (1951). Analysis of production as an efficient combination of activities. In T. Koopmans (Ed.), *Activity Analysis of Production and Distribution* (pp. 33–97). New York: J. Wiley & Sons.
Korhonen, P., & Syrjänen, M. (2004). Resource allocation based on efficiency analysis. *Management Science, 50*, 1134–1144.
Korhonen, P., Tainio, R., & Wallenius, J. (2001). Value efficiency analysis of academic research. *European Journal of Operational Research, 130*, 121–132.
Kuo, R. J., & Lin, Y. J. (2012). Supplier selection using analytic network process and data envelopment analysis. *International Journal of Production Research, 50*, 2852–2863.
Kuosmanen, T., & Kortelainen, M. (2005). Measuring eco-efficiency of production with data envelopment analysis. *Journal of Industrial Ecology, 9*, 59–72.
Lampe, H., & Hilgers, D. (2015). Trajectories of efficiency measurement: A bibliometric analysis of DEA and SFA. *European Journal of Operational Research, 240*, 1–21.
Li, X.-B., & Reeves, G. R. (1999). A multiple criteria approach to data envelopment analysis. *European Journal of Operational Research, 115*, 507–517.
Liu, F.-H. F., Huang, C.-C., & Yen, Y.-L. (2000). Using DEA to obtain efficient solutions for multi-objective 0-1 linear programs. *European Journal of Operational Research, 126*, 51–68.
Liu, J., Lu, L., Lu, W., & Lin, B. (2013a). Data envelopment analysis 1978-2010: A citation-based literature survey. *Omega, 41*, 3–15.
Liu, J., Lu, L., Lu, W., & Lin, B. (2013b). A survey of DEA applications. *Omega, 41*, 893–902.
Liu, W., Meng, W., & Zhang, D. (2010). DEA models with undesirable inputs and outputs. *Annals of Operations Research, 173*, 177–194.
Liu, W., Sharp, J., & Wu, Z. (2006). Preference, production and performance in data envelopment analysis. *Annals of Operations Research, 145*, 105–127.
Lozano, S., Iribarren, D., Moreira, M., & Feijoo, G. (2009). The link between operational efficiency and environmental impacts: A joint application of life cycle assessment and data envelopment analysis. *Science of the Total Environment, 407*, 1744–1754.
Lueddeckens, S., Saling, P., & Guenther, E. (2020). Temporal issues in life cycle assessment: A systematic review. *The International Journal of Life Cycle Assessment*. https://doi.org/10.1007/s11367-020-01757-1.
Madlener, R., Antunes, C. H., & Dias, L. C. (2009). Assessing the performance of biogas plants with multi-criteria and data envelopment analysis. *European Journal of Operational Research, 197*, 1084–1094.
Mahdiloo, M., Saen, R. F., & Lee, K.-H. (2015). Technical, environmental and eco-efficiency measurement for supplier selection: An extension and application of data envelopment analysis. *International Journal of Production Economics, 168*, 279–289.
Malekmohammadi, N., Hosseinzadeh Lotfi, F., & Jaafar, A. B. (2011). Target setting in data envelopment analysis using MOLP. *Applied Mathematical Modelling, 35*, 328–338.
Martin-Gamboa, M., Iribarren, D., Garcia-Gusano, D., & Dufour, J. (2017). A review of life-cycle approaches coupled with data envelopment analysis within multi-criteria decision analysis for sustainability assessment of energy systems. *Journal of Cleaner Production, 150*, 164–174.
Mavrotas, G., & Trifillis, P. (2006). Multicriteria decision analysis with minimum information: Combining DEA and MAVT. *Computers & Operations Research, 33*, 2083–2098.
Miettinen, P., & Hämäläinen, R. P. (1997). How to benefit from decision analysis in environmental life cycle assessment (LCA). *European Journal of Operational Research, 102*, 279–294.
Moons, K., Waeyenbergh, G., & Pintelon, L. (2019). Measuring the logistics performance of internal hospital supply chains: A literature study. *Omega, 82*, 205–217.
Myhre, G., et al. (2013). Anthropogenic and natural radiative forcing. In T. Change Stocker (Ed.), *Climate Change 2013: The Physical Science Basis* (pp. 659–740). Cambridge/New York: Cambridge University Press.
Neely, A., Gregory, M., & Platts, K. (1995). Performance measurement system design: A literature review and research agenda. *International Journal of Operations and Production Management, 15*, 80–110.
Ng, C. Y., & Chuah, K. B. (2014). Evaluation of design alternatives' environmental performance using AHP and ER approaches. *IEEE Systems Journal, 8*, 1182–1189.
Nunamaker, T. R. (1985). Using data envelopment analysis to measure the efficiency of non-profit organizations: A critical evaluation. *Managerial and Decision Economics, 6*, 50–58.
Olesen, O. B., & Petersen, N. C. (2016). Stochastic data envelopment analysis: A review. *European Journal of Operational Research, 251*, 2–21.
Parkan, C., & Wu, M. L. (1997). On the equivalence of operational performance measurement and multiple attribute decision making. *International Journal of Production Research, 35*, 2963–2988.
Podinovski, V. V. (2004). Production trade-offs and weight restrictions in data envelopment analysis. *Journal of the Operational Research Society, 55*, 1311–1322.
Podinovski, V. V. (2015). DEA models with production trade-offs and weight restrictions. In J. Zhu (Ed.), *Data Envelopment Analysis* (pp. 105–144). Boston: Springer.
Qorri, A., Mujkic, Z., & Kraslawski, A. (2018). A conceptual framework for measuring sustainability performance of supply chains. *Journal of Cleaner Production, 189*, 570–584.
Quariguasi Frota Neto, J., Bloemhof-Ruward, J. M., van Nunen, J. A., & van Heck, E. (2008). Designing and evaluating sustainable logistics networks. *International Journal of Production Economics, 111*, 195–208.
Quariguasi Frota Neto, J., Walther, G., Bloemhof, J., Van Nunen, J., & Spengler, T. (2009). A methodology for assessing eco-efficiency in logistics networks. *European Journal of Operational Research, 193*, 670–682.
Roberts, F. S. (1979). *Measurement Theory with Applications to Decision Making, Utility and the Social Sciences*. Reading: Addison-Wesley.
Rubem, A. P., Soares de Mello, J. C., & Angulo Meza, L. (2017). A goal programming approach to solve the multiple criteria DEA model. *European Journal of Operational Research, 260*, 134–139.

Saaty, T. L. (1990). How to make a decision: The Analytic Hierarchy Process. *European Journal of Operational Research, 48*, 9–26.
San Cristobal, J. R. (2011). A multi criteria data envelopment analysis to evaluate the efficiency of the Renewable Energy technologies. *Renewable Energy, 36*, 2742–2746.
Sarkis, J. (2000). A comparative analysis of DEA as a discrete alternative multiple criteria decision tool. *European Journal of Operational Research, 123*, 543–557.
Shen, W., Yang, G., Zhou, Z., & Liu, W. (2019). DEA models with Russel measures. *Annals of Operations Research, 278*, 337–359.
Shephard, R. W. (1970). *Theory of Cost and Production Functions*. Princeton N.J.: Princeton Univ. Press.
Soheilirad, S., Govindan, K., Mardani, A., Zavadskas, E., Nilashi, M., & Zakuan, N. (2018). Application of data envelopment analysis models in supply chain management: A systematic review and meta-analysis. *Annals of Operations Research, 271*, 915–969.
Song, M., An, Q., Zhang, W., Wang, Z., & Wu, J. (2012). Environmental efficiency evaluation based on data envelopment analysis: A review. *Renewable and Sustainable Energy Reviews, 16*, 4465–4469.
Stewart, T. J. (1996). Relationships between data envelopment analysis and multicriteria decision analysis. *Journal of the Operational Research Society, 47*, 654–665.
Sueyoshi, T., Shang, J., & Chiang, W. C. (2009). A decision support framework for internal audit prioritization in a rental car company: A combined use between DEA and AHP. *European Journal of Operational Research, 199*, 219–231.
Tavana, M., Ebrahimnejad, A., Santos-Arteaga, F. J., Mansourzadeh, S. M., & Matin, R. K. (2018). A hybrid DEA-MOLP model for public school assessment and closure decision in the City of Philadelphia. *Socio-Economic Planning Systems, 61*, 70–89.
ten Raa, T. (2019). Performance: The output/input ratio. In T. ten Raa, & W. H. Greene (Eds.), *The Palgrave Handbook of Economic Performance Analysis* (pp. 1–8). Cham: Palgrave MacMillan.
ten Raa, T., & Greene, W. H. (2019). *The Palgrave Handbook of Economic Performance Analysis*. Cham: Palgrave MacMillan.
Thanassoulis, E., & Allen, R. (1998). Simulating weight restrictions in data envelopment analysis by means of unobserved DMUs. *Management Science, 44*, 586–594.
Thies, C., Kieckhaefer, K., Spengler, T. S., & Sodhi, M. S. (2019). Operations Research for sustainability assessment of products: A review. *European Journal of Operational Research, 274*, 1–21.
Vazquez-Rowe, I., Iribarren, D., Moreira, M., & Feijoo, G. (2010). Combined application of life cycle assessment and data envelopment analysis as a methodological approach for the assessment of fisheries. *International Journal of Life Cycle Assessment, 15*, 272–283.
Wallenius, J., Dyer, J., Fishburn, P., Steuer, R., Zionts, S., & Deb, K. (2008). Multiple criteria decison making, multiattribute utility theory: Recent accomplishments and what lies ahead. *Management Science, 54*, 1336–1349.
Weber, M. (1921). *Wirtschaft und Gesellschaft: Grundriss der verstehenden Soziologie* (5th ed.). Tübingen: Mohr Siebeck 1976.
Weber, M., Henderson, A., & Parsons, T. (1964). *The Theory of Social and Economic Organization*. London: Collier-MacMillan.
Wojcik, V., Dyckhoff, H., & Clermont, M. (2019). Is data envelopment analysis a suitable tool for performance measurement and benchmarking in non-production contexts? *Business Research, 12*, 559–595.
Wojcik, V., Dyckhoff, H., & Gutgesell, S. (2017). The desirable input of undesirable factors in Data Envelopment Analysis. *Annals of Operations Research, 259*, 461–484.
Wong, B. Y., Luque, M., & Yang, J.-B. (2009). Using interactive multiobjective methods to solve DEA problems with value judgements. *Computers & Operations Research, 36*, 623–636.
Wong, Y.-H. B., & Beasley, J. E. (1990). Restricting weight flexibility in data envelopment analysis. *Journal of the Operational Research Society, 41*, 829–835.
Yalzin, N., Bayrakdaroglu, A., & Kahraman, C. (2012). Application of fuzzy multi-criteria decision making methods for financial performance evaluation of Turkish manufacturing industries. *Expert Systems with Applications, 39*, 350–364.
Yang, J.-B., Wong, B. Y., Xu, D.-L., & Stewart, T. J. (2009). Integrating DEA-oriented performance assessment and target setting using interactive MOLP methods. *European Journal of Operational Research, 195*, 205–222.
Yang, T. H., & Kuo, C. W. (2003). A hierarchical AHP/DEA methodology for the facilities layout design problem. *European Journal of Operational Research, 147*, 128–136.
Yu, G., Wei, Q., & Brockett, P. (1996). A generalized data envelopment analysis model: A unification and extension of existing methods for efficiency analysis of decision making units. *Annals of Operations Research, 66*, 47–89.
Zhou, P., Ang, B. W., & Poh, K. L. (2008). A survey of data envelopment analysis in energy and environmental studies. *European Journal of Operational Research, 189*, 1–18.
Zhou, Z., & Liu, W. (2015). DEA models with undesirable inputs, intermediates, and outputs. In J. Zhu (Ed.), *Data Envelopment Analysis* (pp. 415–446). New York: Springer.