Ahn, Heinz; Clermont, Marcel; Langner, Julia

The impact of selected input and output factors on measuring research efficiency of university research fields: insights from a purpose-, field-, and method-specific perspective

Journal of Business Economics

Provided in Cooperation with:
Springer Nature

Suggested Citation: Ahn, Heinz; Clermont, Marcel; Langner, Julia (2022) : The impact of selected input and output factors on measuring research efficiency of university research fields: insights from a purpose-, field-, and method-specific perspective, Journal of Business Economics, ISSN 1861-8928, Springer, Berlin, Heidelberg, Vol. 92, Iss. 8, pp. 1303-1335, https://doi.org/10.1007/s11573-022-01099-4

Journal of Business Economics (2022) 92:1303–1335
https://doi.org/10.1007/s11573-022-01099-4

ORIGINAL PAPER

The impact of selected input and output factors on measuring research efficiency of university research fields: insights from a purpose-, field-, and method-specific perspective

Heinz Ahn[1] · **Marcel Clermont**[2] · **Julia Langner**[1]

Accepted: 22 April 2022 / Published online: 8 June 2022

Abstract

The measurement of universities' efficiency is gaining importance due to the increasing competitive pressure. As the results of such a measurement depend especially on the included input and output factors, it is essential to select them carefully to generate valid results. Against this background, we analyze the impact of including selected input and output factors on measuring the research efficiency of university research fields by applying best-practice frontier benchmarking methods. Based on a standard scenario, we consider the additional inclusion of expenditures, research grants, and bibliometric indicators. In doing so, we take account of different research fields as well as methods and purposes of measuring efficiency. Regarding the latter aspect, we distinguish between the computed efficiency degrees and the resulting ranking. We observe that the additional input and output factors have a quite heterogeneous impact on the respective evaluation criteria. The corresponding findings are relatively robust with respect to the research fields and methods considered. Our results provide guidance for evaluators as well as university and political decision-makers.

Keywords Research efficiency of university research fields · Influence of input and output factors · Expenditures · Research grants · Bibliometric indicators · Best-practice frontier benchmarking methods

✉ Heinz Ahn
hw.ahn@tu-braunschweig.de

[1] Department of Business Sciences, Technische Universität Braunschweig, Fallersleber-Tor-Wall 23, 38100 Brunswick, Germany

[2] Cooperative University Gera-Eisenach, Am Wartenberg 2, 99817 Eisenach, Germany

1 Introduction

The amount of funding allocated to universities and their organizational units is increasingly based on multidimensional performance indicators, which are, for example, the subject of target agreements (e.g., De Fraja et al. 2019). To compare the considered units, it is necessary to aggregate these performance indicators into one-dimensional measures, such as effectiveness and efficiency degrees. The selection of performance indicators used to quantify university performance has a decisive influence on these degrees. In the following, we will deal with this aspect of indicator selection, whereby we focus on measuring the research efficiency of university research fields.

A special aggregation method is necessary to determine research efficiency since relevant university performance indicators cannot be aggregated to financial ratios due to the lack of market prices. For this purpose, empirical studies often use methods based on production theory. The advantage of these methods is that the evaluators do not have to specify weights for the performance indicators since the methods do so endogenously.

The production-theoretical relation is expressed on the one hand through the fact that the performance indicators usually are categorized as input factors to be minimized and output factors to be maximized. On the other hand, the methods determine a best-practice frontier in the form of a production function based on the input-output data. The distance between this function and the data points determines the resulting efficiency degree. Corresponding methods differ in how the distance between the production function and a data point is interpreted: deterministic methods interpret it completely as inefficiency, whereas stochastic methods assume that a part of the distance is caused by stochastic noise.

Concerning deterministic methods, the Data Envelopment Analysis (DEA) established by Charnes et al. (1978) is predominantly applied in the literature for efficiency measurement in the university context. Furthermore, the deterministic Normalized Additive Analysis (NAA)[1] introduced by Ahn et al. (2007) also seems to be an interesting option. Its advantage is the easy calculation of efficiency degrees, making it particularly suitable for practical use without expert support.

Regarding stochastic methods, the Stochastic Frontier Analysis (SFA) introduced by Aigner et al. (1977) and Meeusen and Van den Broeck (1977) predominantly attracts attention in the literature. An interesting alternative is the Stochastic Non-Smooth Envelopment of Data (StoNED) developed by Kuosmanen and Kortelainen (2012), although, to our knowledge, it has not yet been used in a university context. It combines ideas from DEA and SFA and thus may retain their advantages without significant disadvantages.

Various studies show that the results of efficiency analyses can strongly depend on the choice of method (e.g., Andor and Hesse 2014). Therefore, parallel analyses

[1] This designation corresponds to that of Ahn et al. (2012). Büschken (2009) used a similar approach, which he calls naïve efficiency measurement.

with varying methods are appropriate. Accordingly, we compare the methods mentioned above in our study.

Although the selection of input and output factors also obviously influences the efficiency degree, their selection rarely seems systematic. To date, no standard has been established for input and output factors to determine research efficiency in the university context. It is important to emphasize that these factors should be selected sparingly since too many factors can lead to very high, hardly discriminating efficiency degrees when the methods mentioned above are applied (Dyson et al. 2001).

Therefore, our objective is to investigate the influence of the inclusion of selected input and output factors on the efficiency measurement results, depending on the purpose of the analysis, the analyzed research field, and the efficiency measurement method. Based on a standard scenario with common input and output factors, we analyze the extent to which the extension of this scenario by expenditures, research grants, and bibliometric indicators leads to a shift in the efficiency measurement results. The resulting contributions of our study can be summarized as follows:

1. We provide guidance for the choice of input and output factors to measure the efficiency of university research.
2. Among other things, to cover the dissemination and reception of research results, we include four citation measures that have not yet been evaluated in direct comparison with each other: citations, citations per publication, h-index, and J-factor.
3. We differentiate between two purposes of analysis, the accurate determination of the ranking and the accurate estimation of the efficiency degrees. It is shown that the two purposes react with different sensitivity to additional input/output factors.
4. We consider six research fields at German universities to account for field-specific differences, e.g., publication habits.
5. In addition to DEA and SFA, for the first time, we include StoNED and NAA in analyses of the impact of the selected input and output factors in the university context.

Our paper is structured as follows: Sect. 2 deduces the state of the art and research gaps. Section 3 presents the study's research design. In Sects. 4, 5 and 6, we analyze the impact of including the additional input and output factors in our standard scenario. Section 7 discusses the results of this study. The paper concludes with a critical evaluation of the study and an outlook in Sect. 8.

2 State of the art and research gaps

We used a systematic literature search to examine which studies have been undertaken to analyze the influence of additional input and output factors on the efficiency degree of university research or their research fields. 16 publications showed to be relevant to our study. For each of these, Table 1 shows the input and output factors that were used in all scenarios analyzed as well as the input and output factors

Table 1 Input and output factors used in the relevant publications from the literature review

	Factors used in all scenarios	Varied factors
Johnes and Johnes (1992)	O: publications (by type)	I: staff (by task), research grants per capita
Johnes and Johnes (1993)	I: staff (by task)	I: research grants per capita O: publications (by type), research grants
Johnes (1995)		I: staff (by funding) O:publications (by type), research grants
Johnes and Johnes (1995)	O: publications (by type)	I: staff (by task), research grants per capita, staff's time available for research
McMillan and Datta (1998)		I: staff (science/non-science subject category), expenditures (by type) O: research grants, share of research grants (by origin), students (by degree sought and science/non-science subject category)
Johnes (2006)		I: staff, expenditures (by type), students (by degree sought) O: research grants, graduates (by degree)
Johnes and Yu (2008)	I: percentage of staff with associate professor position or higher, research expenditures, library books index, buildings' area index O: publications per academic, reputation index	I: staff to student ratio, percentage of post graduate students O: publications
Rayeni and Saljooghi (2010)		I: staff, guest lecturers, students O: performed research work, graduates, passed students to higher levels
Kounetas et al. (2011)	I: staff, expenditures	I: graduate students, postgraduate students O: publications, conferences
Agasisti et al. (2012)	I: staff, laboratories	O: publications, citations per publication, h-index, research grants (by origin), revenues from orders
Bielecki and Albers (2012)	I: professors, research funds per professor, doctorates per professor	I: habilitations per professor O: weighted national visible publications, internationally visible publications
De la Torre et al. (2017)	I: staff, students O: publications, graduates	O: intellectual or industrial property agreements, income that the university can attract through its scientific work, university spin-offs

Table 1 (continued)

	Factors used in all scenarios	Varied factors
Expósito-García and Velasco-Morente (2018)	I: expenditures per student, expenditures per academic, research periods per academic O: non-cited publications (undesirable output)	O: publications per 100 academics, publications in Q1 journals per 100 academics
Gralka et al. (2019)	I: expenditures, wages O: graduates (science/non-science subject category)	O: publications, 10% most-cited publications, 1% most-cited publications, research grants
Geissler et al. (2021)	I: expenditures O: students (by research field)	O: publications, research grants, Mean Normalized Citation Score, Category Normalized Citation Impact, Journal Normalized Citation Impact
Ghimire et al. (2021)		I: staff, expenditures O: publications, research grants, students, student's satisfaction level

I inputs, *O* outputs

that were varied throughout the respective analysis. The publications' results that are interesting in our context are briefly presented below.

Most of the mentioned studies used DEA as method for measuring efficiency. Johnes and Johnes (1992, 1993), and Johnes (1995) compared respective efficiency degrees for departments of economics in the U.K. Mostly varying more than one input or output factor, their results changed with different intensity depending on the variant. Based on these studies, Johnes and Johnes (1995) changed the research setting slightly and found a quite high correlation of 91.9% between the estimated efficiency degrees for the different variants. Similarly, McMillan and Datta (1998) found moderate deviations when they investigated the influence of nine different input-output combinations on the estimated efficiency degrees of Canadian universities. While McMillan and Datta varied at least two input or output factors at a time, Johnes (2006) analyzed the effects of excluding one to three of the input and output factors on the efficiency degrees of British universities. It was found that the exclusion of staff and expenditures on library and information services had almost no effect on the estimated efficiency degrees, and the rank correlation between the results was 92%. This effect was not observed for the other factors. Johnes and Yu (2008) investigated the efficiency of Chinese universities. Depending on the input factors considered, rank correlations between the efficiency degrees with and without including the publications were at least 98.2%.

Rayeni and Saljooghi (2010) examined Iranian universities by excluding one input or output factor in each scenario. They found out that the exclusion of staff led to the biggest changes in the efficiency degrees and a correlation of 18.3% while the other variations resulted in clearly smaller changes and higher correlations. The analysis of the departments of a Greek university by Kounetas et al. (2011) showed that the changes in efficiency degrees varied depending on the variant under consideration, especially when excluding publications. Agasisti et al. (2012) simultaneously varied several output factors for Italian science departments. Depending on the variant, quite different efficiency degrees with rather low correlations between 3.9 and 55.8% were found. Again, several output factors were varied simultaneously, but the effect of a single factor was not considered.

Bielecki and Albers (2012) examined German business schools. Big changes in efficiency degrees emerged, e.g., weighted nationally visible publications increased the efficiency degrees. The effect of habilitations per professor, for which different years were taken as a basis, could not be determined. The rank correlations were low with a maximum of 26%, often even negative, but rarely significant. The efficiency of Spanish universities was investigated by De la Torre et al. (2017). They found rank correlations between 70 and 97%. Furthermore, concerning Spanish universities, Expósito-García and Velasco-Morente (2018) observed minimal differences in efficiency degrees when varying publications as an output factor.

In addition to DEA, Gralka et al. (2019) used SFA to analyze the efficiency of German universities. The average efficiency degrees for the output variations were similar, although the deviations were somewhat higher for SFA. Compared to those with research grants, the rank correlations for the different analyses ranged from 84.9 to 86.3% for DEA and from 68.5 to 83.3% for SFA. Also comparing German

universities by applying SFA, Geissler et al. (2021) investigated the impact of varying the output factor on the estimated efficiency degrees. They found that the consideration of normalized measures led to a significant reduction in the average efficiency degree and stronger discrimination of the top group. Comparing the three normalized measures did not yield any considerable changes concerning both efficiency degrees and ranking. Finally, Ghimire et al. (2021) developed a DEA model to evaluate the efficiency of Canadian universities by using diverse weights to model the variation of input and output factors. Their study showed a crucial role of the selection of inputs and outputs in determining the efficiencies and rankings.

Regarding the state of the art, various research gaps can be identified concerning the research questions outlined in the introduction. First, DEA was the dominant aggregation method used in the identified studies. SFA was used in only two studies. Novel methods, such as StoNED and NAA, have not been discussed. Thus, to date, there is no comprehensive study that compares the inclusion of input and output factors concerning the above-mentioned efficiency measurement methods.

Furthermore, there lacks systematic analysis of the impact of including single input and output factors, as several factors are often varied simultaneously. Such approaches have the disadvantage that the strength of the respective influence of the individual input or output factors is not visible. Moreover, such effects were not examined by comparing different research fields at once since the mentioned studies only analyzed single departments or universities as a whole.

Concerning the possible purposes of a performance analysis, the studies mainly compared the estimated efficiency degrees. Such a comparison is only of limited usefulness due to method-immanent changes in the average level of efficiency degrees for different numbers of considered input and output factors (see Sect. 3.2). Moreover, the purpose of accurately determining the ranking has been addressed in only a few publications.

3 Research design

Based on the identified research gaps, we present our research design: Sect. 3.1 explains which specifications we consider for the respective methods to measure efficiency. Section 3.2 presents the evaluation criteria used for the analyses. In Sect. 3.3, we describe the underlying standard scenario and data set.

3.1 Model specification

As stated in the introduction, we use DEA, NAA, SFA, and StoNED to measure efficiency. Different models have been developed for each of these methods. We choose those models that are predominantly applied in empirical applications. In general, we assume an output orientation since universities can influence outputs rather than inputs in the short term.

For DEA, we use the two basic models DEA_{BCC} (Banker et al. 1984) and DEA_{CCR} (Charnes et al. 1978); the former assumes variable, the latter constant

returns to scale. We use the model with arithmetic mean for NAA, according to Ahn et al. (2007). SFA is used under the assumption of a Cobb-Douglas function. Since it includes only one output factor by default, ray production functions are used to model multiple output factors (see for SFA, e.g., Löthgren 1997, and for StoNED, e.g., Schaefer and Clermont 2018). According to Schaefer and Clermont (2018), these have an advantage over distance functions; however, the limitations addressed by Henningsen et al. (2017) have to be taken into account. To circumvent some of these limitations, the angles for SFA are exponentiated before including them in the function, as suggested by the authors. Owing to the theoretically and empirically established dominance of the maximum likelihood method over the method of moments for the partitioning of the error term, the maximum likelihood method is used (Olson et al. 1980, for an empirical comparison, see, e.g., Andor and Hesse 2014). The error term is modeled multiplicatively. A half-normal distribution is assumed concerning the inefficiency term, and with respect to the noise term, we assume a normal distribution. The Battese-Coelli estimator is chosen as point estimator (Battese and Coelli 1988, p. 390) since it minimizes the mean squared error compared to the estimators proposed by Jondrow et al. (1982) (Bogetoft and Otto 2011). We use the described specifications for SFA also for StoNED. Only the partitioning of the error term is, due to the model, not done with the maximum likelihood method, but with the pseudo-likelihood method (Kuosmanen and Kortelainen 2012).

Since the stochastic methods are not unit invariant when using ray production functions, all empirical data presented below are normalized based on their arithmetic mean, following the proposal of Henningsen et al. (2017). That way, dependencies on the unit are excluded (Dyson et al. 2001).

3.2 Evaluation criteria

Efficiency measurement serves mainly two purposes: accurate determination of the ranking and accurate estimation of the efficiency degrees. To evaluate the impact of the input and output factors on the respective efficiency measurement results, quantitative criteria are required that allow statements on the similarity of the results. Thus, with respect to ranking, the similarity between ranking positions is of interest. In the following, we evaluate the extent to which the rankings resulting from different input and output factors coincide by using the rank correlation (RC) according to Spearman (1904). It is determined by

$$RC = \frac{Cov\left(r\left(TE_1\right), r\left(TE_2\right)\right)}{\sqrt{Var\left(r\left(TE_1\right)\right) Var\left(r\left(TE_2\right)\right)}}, \tag{1}$$

where $r\left(TE_1\right)$ and $r\left(TE_2\right)$ represent the corresponding rankings of the efficiency degrees for two analyses with different input or output factors.

Suppose, conversely, the focus is on estimating the efficiency degrees as accurately as possible. In that case, it should be noted that increasing the number of input

and output factors considered can already inherently lead to an increase in the estimated efficiency degrees (for DEA, see, e.g., Sexton et al. 1986, and Dyson et al. 2001). Due to this effect, considering the change in efficiency degrees themselves, for example, via the mean absolute deviation, does not seem reasonable. To level out this effect, we consider the standard deviation of mean deviation (SDMD), which provides information about the spread of the changes in the efficiency degrees. It is calculated by

$$SDMD = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(\left(TE_{1,j} - TE_{2,j}\right) - \frac{1}{n}\sum_{j=1}^{n}\left(TE_{1,j} - TE_{2,j}\right)\right)^2}, \tag{2}$$

where $TE_{1,j}$ and $TE_{2,j}$ represent the efficiency degree of university j for two analyses with varying input or output factors, and n indicates the number of universities.

Assuming that, for example, taking additional input or output factors into account would lead to a shift of all efficiency degrees by five percentage points. This would result in an SDMD of 0. If, conversely, there are large differences in the shifts of efficiency degrees, this would lead to a high SDMD.

3.3 Standard scenario and selection of data

To analyze the impact of selected input and output factors on measuring research efficiency, we use a standard scenario as a benchmark. In this regard, we consider two input factors and one output factor frequently applied in studies (see also the reviews by Berbegal-Mirabent and Solé Parellada 2012; De Witte and López-Torres 2017; Gralka 2018). As staff-related input factors, we use the number of professors and the number of research assistants. The data were obtained from the German Federal Statistical Office in a special evaluation. This institution subdivided the data concerning German universities into 89 areas of research, which we grouped into eleven research fields.

As output factor of the standard scenario, we use the number of journal publications to indicate the generation of public knowledge. The publication data were collected using the Web of Science database hosted by the Competence Centre for Bibliometrics,[2] enhanced with institution encoding. The publications were allocated to the eleven research fields based on the journal in which the article was published. Thereby, we draw on the science classification system of Archambault et al. (2011) to avoid overlap in the assignment of journals to research fields. An extension of the original classification system was used that considers some journals not included before (Tunger et al. 2022). This classification system allocates each journal to precisely one subfield, which we aggregated to our research fields.

Due to the big differences in the publication and citation habits, as well as in the available resources of different research fields (Linton et al. 2011), we conduct

[2] Details on the Competence Centre for Bibliometrics are available at http://www.bibliometrie.info/en.

Table 2 Descriptive statistics on the standard scenario in 2010

Research field (# of data sets)	Professors	Research assistants	Publications	Research field (# of data sets)	Professors	Research assistants	Publications
B&E (74)				Engineering (40)			
Min	7	5	1	Min	2	4	4
Max	68	313	133	Max	144	2,087	570
Mean	22.2	89.4	22.1	Mean	59.2	518.2	148.3
Median	19.0	73.0	16.0	Median	51.0	355.0	108.0
Std.dev	12.3	63.2	21.1	Std.dev	38.6	504.9	130.9
Humanities (63)				Math. & CS (70)			
Min	6	18	1	Min	6	5	4
Max	166	494	45	Max	66	512	202
Mean	58.9	149.1	13.6	Mean	29.2	128.2	60.2
Median	52.0	117.0	8.0	Median	28.0	100.0	47.5
Std.dev	36.4	113.9	12.2	Std.dev	13.4	100.8	43.5
Natural Sc (66)				Social Sc (72)			
Min	4	6	4	Min	2	2	2
Max	147	1,082	1,471	Max	90	327	166
Mean	58.8	403.3	510.9	Mean	30.0	106.8	46.7
Median	56.0	392.0	478.5	Median	28.0	89.0	36.0
Std.dev	34.7	274.6	372.0	Std.dev	18.2	72.3	38.0

B&E Business & Economics, *Math. & CS* Mathematics & Computer Science, *Natural Sc* Natural Sciences, *Social Sc* Social Sciences

separate investigations by research field. From the eleven research fields mentioned above, we selected the following six for our analyses based on data availability and sufficient publication activity: business & economics, engineering, humanities, mathematics & computer science, natural sciences, and social sciences.

For the data collection, the year 2010 was chosen to ensure a sufficiently long citation period for the bibliometric indicators (Clermont et al. 2021). It should be noted that the focus here is not on empirical statements about a specific year and individual universities but on conceptual analyses. A plausibility analysis was conducted for the data sets of the standard scenario and the other input and output factors so that only complete and consistent data sets were included in the analyses.

Table 2 shows the descriptive data of the standard scenario for the six research fields investigated. Depending on the research field, between 40 and 74 universities are included in the analyses. The differences in habits between the research fields can be seen. Specifically, engineering and natural sciences are characterized by relatively high numbers of professors, research assistants, and publications.

To briefly give a general impression, it is observed that for the standard scenario, the efficiency degrees for NAA are clearly lower than for the other methods. Additionally,

Table 3 Considered inputs and outputs for variations of expenditures

Variants concerning exp	Professors	Research asst	Publications	Pers. exp	Eq. exp	Total exp
– (std.sc)	I	I	O			
Pers. exp	I	I	O	I		
Eq. exp	I	I	O		I	
Pers. & eq. exp	I	I	O	I	I	
Total exp (= sum of pers. & eq. exp)	I	I	O			I

asst assistants, *eq*. equipment, *exp* expenditures, *I* input, *O* output, *pers*. personnel, *std.sc* standard scenario

most of the time, for the deterministic methods (DEA, NAA) the median of the efficiency degrees is lower than the mean. That indicates that the corresponding distribution of estimated efficiency degrees is right-skewed. For the stochastic methods (SFA, StoNED), the opposite is true. Regarding the research fields, no general difference is visible.

4 Influence of expenditures

In addition to staff, universities also spend financial resources on research, which is another input factor to be minimized. In contrast to studies that exclusively consider staff (e.g., Lee and Worthington 2016; Wohlrabe and Friedrich 2017; Ibrahim and Fadhli 2021), some studies include financial resources in addition to staff (e.g., Wang and Hu 2017; Duan 2019; Jiang et al. 2020).

In the following, we examine how the efficiency degrees change if we integrate expenditures as an additional input factor in the standard scenario. In our data set, the research fields' expenditures are divided into equipment[3] and personnel expenditures. In the latter case, it can be assumed that these are already to some extent covered by the input factors for staff (professors and research assistants) so that we are aware of the issue of double counting. Therefore, we differentiate according to the type of expenditures and analyze five variants of taking expenditures into account, as listed in Table 3.

4.1 Descriptive analysis

Table 4 presents the descriptive statistics for the expenditure data, which are as well extracted from the special evaluation of the German Federal Statistical Office. Again, there are clear differences regarding the research fields, with the highest expenditures being made in engineering and natural sciences. The last rows of

[3] Thereby, expenditures on equipment summarize expenditures that are not spent for staff, e.g., expenditures on material, books, machines, rent, energy, interest etc.

Table 4 Descriptive statistics on expenditures in 2010

		Pers. exp (in mio. €)	Eq. exp (in mio. €)	Total exp (in mio. €)			Pers. exp (in mio. €)	Eq. exp (in mio. €)	Total exp (in mio. €)
B&E	Min	1.2	0.2	1.5	Engineering	Min	1.1	0.2	1.5
	Max	23.5	33.6	56.2		Max	195.9	66.5	262.4
	Mean	7.4	2.4	9.8		Mean	45.2	18.4	63.6
	Median	5.9	1.0	6.9		Median	30.7	12.2	43.2
	Std.dev	4.9	4.7	8.7		Std.dev	43.7	16.8	59.7
	Ratio of total exp	0.751	0.249			Ratio of total exp	0.711	0.289	
Humanities	Min	2.2	0.2	2.7	Math. & CS	Min	1.2	0.1	1.3
	Max	44.0	27.6	63.2		Max	40.8	17.9	52.0
	Mean	16.1	3.8	19.8		Mean	10.8	3.1	13.9
	Median	12.4	1.7	14.1		Median	8.5	1.7	10.7
	Std.dev	11.4	5.1	15.4		Std.dev	8.1	3.2	10.7
	Ratio of total exp	0.809	0.191			Ratio of total exp	0.776	0.224	
Natural Sc	Min	0.8	0.3	1.1	Social Sc	Min	0.5	0.1	0.5
	Max	77.0	63.3	129.3		Max	29.3	11.7	38.3
	Mean	29.4	16.8	46.1		Mean	8.9	2.2	11.1
	Median	29.2	12.7	43.7		Median	6.8	1.6	9.3
	Std.dev	19.7	14.6	33.1		Std.dev	6.0	2.2	7.7
	Ratio of total exp	0.636	0.364			Ratio of total exp	0.799	0.201	

Table 5 Pearson correlations between expenditures and staff

		Professors	Research asst	Pers. exp	Eq. exp	Total exp			Professors	Research asst	Pers. exp	Eq. exp	Total exp
B&E	Professors	1					Engineering	Professors	1				
	Research asst	.778	1					Research asst	.870	1			
	Pers. exp	.850	.773	1				Pers. exp	.867	.986	1		
	Eq. exp	.404	*.161*	.666	1			Eq. exp	.805	.939	.935	1	
	Total exp	.692	.519	.916	.909	1		Total exp	.861	.986	.995	.966	1
Humanities	Professors	1					Math. & CS	Professors	1				
	Research asst	.923	1					Research asst	.868	1			
	Pers. exp	.940	.941	1				Pers. exp	.882	.969	1		
	Eq. exp	.561	.670	.710	1			Eq. exp	.662	.634	.725	1	
	Total exp	.879	.916	.972	.855	1		Total exp	.869	.927	.978	.853	1
Natural Sc	Professors	1					Social Sc	Professors	1				
	Research asst	.942	1					Research asst	.830	1			
	Pers. exp	.938	.965	1				Pers. exp	.903	.928	1		
	Eq. exp	.761	.839	.856	1			Eq. exp	.650	.689	.736	1	
	Total exp	.894	.945	.974	.952	1		Total exp	.883	.913	.982	.852	1

Almost all results are significant with $p \leq 0.05$. Non-significant results are written in italics

Gray-colored cells refer to correlations ≥ 90%

Table 4 show the average share of personnel and equipment expenditures in total expenditure. As expected, personnel expenditures account for the largest share of expenditures in all research fields. Engineering and natural sciences also exhibit a high share of equipment expenditures and, thus, a higher relevance of these expenditures. This could be because research projects in these research fields tend to require more technical equipment.

Table 5 presents the Pearson correlations among the input factors considered. Very high correlations of 90% or more are highlighted in gray here and in all subsequent tables. As can be seen from Table 5, the correlations are mostly above 80% and often even above 90%. The correlations for equipment expenditures are often the lowest, whereas the correlations between staff (especially research assistants), personnel, and total expenditures are very high. This confirms the assumption that personnel expenditures are widely described by staff. When comparing research fields, it is noticeable that the correlations are particularly high for engineering, whereas on average, they are the lowest for business & economics. In addition to the question of how much the results change depending on the factors included, the question arises as to whether these changes are related to the correlations in the initial data.

Table 6 Rank correlations for DEA_{BCC} when including expenditures

		– (std.sc.)	Pers. exp	Eq. exp	Pers. & eq. exp	Total exp			– (std.sc.)	Pers. exp	Eq. exp	Pers. & eq. exp	Total exp
B&E	– (std.sc.)	1					Engineering	– (std.sc.)	1				
	Pers. exp	.980	1					Pers. exp	.999	1			
	Eq. exp	.942	.967	1				Eq. exp	.989	.989	1		
	Pers. & eq. exp	.936	.973	.996	1			Pers. & eq. exp	.989	.989	.999	1	
	Total exp	.975	.996	.982	.984	1		Total exp	.997	.998	.993	.993	1
Humanities	– (std.sc.)	1					Math. & CS	– (std.sc.)	1				
	Pers. exp	.996	1					Pers. exp	.966	1			
	Eq. exp	.934	.935	1				Eq. exp	.962	.979	1		
	Pers. & eq. exp	.933	.938	.999	1			Pers. & eq. exp	.947	.992	.988	1	
	Total exp	.994	.998	.944	.945	1		Total exp	.968	.997	.986	.992	1
Natural Sc	– (std.sc.)	1					Social Sc	– (std.sc.)	1				
	Pers. exp	.935	1					Pers. exp	.997	1			
	Eq. exp	.913	.958	1				Eq. exp	.985	.989	1		
	Pers. & eq. exp	.908	.965	.997	1			Pers. & eq. exp	.984	.990	.999	1	
	Total exp	.928	.988	.984	.984	1		Total exp	.998	.999	.992	.992	1

All results are significant with $p \leq 0.05$

Gray-colored cells refer to correlations $\geq 90\%$

4.2 Results concerning the purpose "determination of the ranking"

We consider the influence of including the various expenditures as an additional input factor on the determination of the ranking based on the RC. Since the results of the methods show the same basic trend, they are presented in the following only for DEA_{BCC} as an example. The corresponding RCs between the results of the different variants are presented in Table 6.

All research fields show very high RCs of (mostly clearly) more than 90%. Especially engineering stands out with RCs of at least 98.9%. This is analogous to the correlations of the initial data. Overall, only marginal differences can be observed between research fields. Furthermore, the RCs are only slightly lower for the variants where the equipment expenditures vary, although these have the lowest correlations in the initial data. Thus, there is no direct relationship between the correlations of the input data in Table 5 and the RCs in Table 6.

Overall, the very high RCs show that the inclusion of the various expenditures has little to no impact on the resulting rankings from DEA_{BCC}. As indicated earlier, this finding also applies to the other methods. Since the resulting rankings hardly change, the expenditures can thus be neglected with regard to this purpose.

Table 7 Standard deviations of mean deviation for DEA_{BCC} and SFA when including expenditures

	DEA_{BCC}* \ SFA*	– (std.sc.)	Pers. exp	Eq. exp	Pers. & eq. exp	Total exp
B&E	– (std.sc.)		.007	.007	.007	.006
	Pers. exp	.062		.006	.002	.002
	Eq. exp	.143	.103		.004	.004
	Pers. & eq. exp	.149	.106	.023		.001
	Total exp	.082	.032	.077	.082	
Humanities	– (std.sc.)		.017	.013	.053	.005
	Pers. exp	.021		.021	.045	.013
	Eq. exp	.099	.099		.046	.016
	Pers. & eq. exp	.100	.098	.013		.052
	Total exp	.026	.012	.097	.096	
Natural Sc	– (std.sc.)		.062	.044	.062	.062
	Pers. exp	.061		.050	.006	.036
	Eq. exp	.076	.056		.055	.026
	Pers. & eq. exp	.076	.048	.019		.042
	Total exp	.063	.025	.036	.033	

	DEA_{BCC}* \ SFA*	– (std.sc.)	Pers. exp	Eq. exp	Pers. & eq. exp	Total exp
Engineering	– (std.sc.)		.019	.010	.032	.017
	Pers. exp	.009		.014	.024	.007
	Eq. exp	.046	.046		.034	.008
	Pers. & eq. exp	.046	.046	.003		.030
	Total exp	.031	.030	.035	.035	
Math. & CS	– (std.sc.)		.011	.005	.017	.007
	Pers. exp	.066		.014	.012	.006
	Eq. exp	.078	.047		.017	.011
	Pers. & eq. exp	.090	.033	.035		.016
	Total exp	.070	.022	.029	.027	
Social Sc	– (std.sc.)		.075	.104	.101	.099
	Pers. exp	.018		.072	.054	.045
	Eq. exp	.056	.052		.024	.037
	Pers. & eq. exp	.058	.048	.016		.015
	Total exp	.017	.011	.043	.042	

*The results in the respective lower triangulars correspond to DEA_{BCC}, whereas the results in the respective upper triangulars correspond to SFA

Gray-colored cells refer to SDMDs ≤ 0.05

4.3 Results concerning the purpose "estimation of efficiency degrees"

We now analyze the influence of expenditures on the estimated efficiency degrees using the SDMD. In contrast to the RC, a stronger dependence on the applied method is evident. In particular, the results of DEA_{BCC} and DEA_{CCR} differ from those of SFA, StoNED, and NAA, which is why we present the results for DEA_{BCC} and SFA as examples in the following. Table 7 presents the corresponding SDMDs of the different variants. Here and in all the following tables, low SDMDs not greater than five percentage points are marked in gray.

For DEA_{BCC}, in contrast to RCs, bigger differences can be observed between the variants and research fields. Since the SDMDs amount up to 14.9 percentage points, there are sometimes clear shifts in the ratios of the efficiency degrees to one another. This effect is particularly evident for business & economics and humanities, whereas the SDMDs are rather low for engineering. Additionally, the spread tends to be particularly high when equipment expenditures are taken into account.

When applying SFA, the resulting changes are considerably smaller than those with DEA_{BCC}. For example, the SDMDs for business & economics, engineering, and mathematics & computer science are always less than five percentage points. For StoNED and NAA, this is also partly the case for other research fields, e.g.,

Table 8 Considered inputs and outputs for variations of grants

Variants concerning grants	Professors	Research asst	Publications	Grants
– (std.sc.)	I	I	O	
Grants as input	I	I	O	I
Grants as output	I	I	O	O

humanities and natural sciences. Overall, the highest spread tends to be found again for the inclusion of equipment expenditures.

In summary, in contrast to the accurate determination of the ranking, accurately estimating the efficiency degrees shows a stronger impact of the expenditures. The absolute values of the SDMDs differ more strongly between the methods and the research fields considered. However, there are no differences in the basic statements, e.g., the highest SDMDs are generally found for the additional inclusion of equipment expenditures. Accordingly, including expenditures, especially equipment expenditures, should be examined more precisely for the addressed purpose, especially when using DEA_{BCC} or DEA_{CCR}.

5 Influence of research grants

An aspect much discussed in research evaluations concerns the extent to which grants should be considered an input factor of research to be minimized or an output factor to be maximized. Generally, the input and output factors selected for a specific efficiency analysis should be deduced from the goals pursued (Ahn and Le 2015; Dyckhoff 2018). From such a goal-oriented view, contradictions between the use of a parameter as an input or output factor can be explained, e.g., depending on the stakeholders involved. For example, a grant donor may pursue the goal of using these grants as economically as possible, which supports its consideration as an input. This corresponds with the production-theoretical point of view, from which grants represent a special form of financial resources. On the other hand, universities aim to increase their performance, competitiveness, and reputation, which can be, e.g., measured by grants since their acquisition largely depends on previous outstanding research (Johnes 1997). From this perspective, grants are interpreted as an output factor.

In addition to various studies that do not consider grants (e.g., Abramo and D'Angelo 2009; Zhang et al. 2016; Ibrahim and Fadhli 2021), some include it as an input factor (e.g., Furková 2013; Jauhar et al. 2018; Wang 2019), whereas others include it as an output factor (e.g., Johnes 1995; Clermont 2016; Gralka et al. 2019). Besides, some authors discuss models with dual-role factors, e.g. grants can be considered as input and output factor simultaneously (e.g., Beasley 1990; Cook et al. 2006). However, in order to ensure consistent investigations across all methods and sections, we only consider standard models and compare the variants listed in Table 8 to analyze the effects of considering grants as input or output.

Table 9 Descriptive statistics on grants (in mio. €) in 2010

	B&E	Engineering	Humanities	Math. & CS	Natural Sc	Social Sc
Min	0.1	0.3	0.1	0.1	0.1	0.1
Max	9.3	139.6	21.2	38.4	48.1	16.2
Mean	1.8	28.5	3.6	5.0	16.6	2.6
Median	1.3	14.7	2.2	3.1	13.7	1.9
Std.dev	1.9	31.3	4.0	6.0	13.1	2.6
Ratio of total exp	0.181	0.449	0.181	0.363	0.360	0.234

Table 10 Pearson correlations between grants, staff, and publications

	B&E	Engineering	Humanities	Math. & CS	Natural Sc	Social Sc
Professors	.546	.812	.782	.717	.897	.694
Research asst	.682	.973	.828	.897	.948	.849
Publications	.584	.830	.794	.703	.924	.632

All results are significant with $p \leq 0.05$

Gray-colored cells refer to correlations $\geq 90\%$

5.1 Descriptive analysis

First, we look at the descriptive statistics in Table 9, where the data are again extracted from the special evaluation of the German Federal Statistical Office. As before, the highest values are found for engineering and natural sciences. If we look at the share of grants in total expenditures, we see that mathematics & computer science also have a very high share and thus a high relevance of grants. The other three research fields differ, with a maximum share of 24%. This is basically in line with expectations since grant-financed projects with high volumes often finance technical equipment, among other things.

The Pearson correlations between grants and the input/output factors of the standard scenario are presented in Table 10. There are quite clear differences visible for the various research fields. For example, engineering and natural sciences show the highest correlations with at least 80% and for some variants, even more than 90%, whereas those for business & economics are the lowest, with a maximum of 68.2%. Furthermore, the correlations between grants and research assistants are the highest across all research fields. Regarding the correlation between grants and publications, natural sciences stand out from the other research fields with 92.4%.

Table 11 Rank correlations for DEA_{BCC} when including grants

		– (std.sc)	Grants as input	Grants as output			– (std.sc)	Grants as input	Grants as output
B&E	– (std.sc)	1			Engineering	– (std.sc)	1		
	Grants as input	.849	1			Grants as input	.989	1	
	Grants as output	.735	.475	1		Grants as output	.774	.705	1
Humanities	– (std.sc)	1			Math. & CS	– (std.sc)	1		
	Grants as input	.911	1			Grants as input	.960	1	
	Grants as output	.898	.767	1		Grants as output	.830	.746	1
Natural Sc	– (std.sc)	1			Social Sc	– (std.sc)	1		
	Grants as input	.914	1			Grants as input	.972	1	
	Grants as output	.769	.536	1		Grants as output	.917	.852	1

All results are significant with $p \leq 0.05$

Gray-colored cells refer to correlations $\geq 90\%$

5.2 Results concerning the purpose "determination of the ranking"

Regarding the RCs, the methods show very similar results, which is why, in analogy to Sect. 4, they are again presented in Table 11 for DEA_{BCC} as an example.

For all research fields, comparing the standard scenario with the inclusion of grants as input leads to the highest RCs. Except for business & economics, these values are above 90%. Thus, considering grants as input has little effect on the ranking. When grants are included as output, the results differ more, with RCs between 73.5% and 91.7%. Despite the high correlation between grants and publications in the initial data, e.g., the RC for natural sciences is only 76.9%. This confirms that a high correlation of the factors does not necessarily imply a high RC of the results. Comparing grants as input with grants as output shows considerably lower correlations with RCs starting at 47.5%, which is expected due to the opposite view. Concerning the research fields, social sciences show the highest RCs on average, whereas business & economics show the lowest RCs.

For the other research fields and methods, the RCs exhibit the same basic trend, and the same findings can be obtained. Thus, overall, it can be concluded that there is a notable impact of the inclusion of grants as an output factor on the ranking. Meanwhile, considering grants as an input factor is less relevant since the resulting rankings hardly change in the presence of RCs of mostly over 90%.

Table 12 Standard deviations of mean deviation for DEA_{BCC} and StoNED when including grants

	DEA_{BCC}* \ StoNED*	– (std.sc)	Grants as input	Grants as output		DEA_{BCC}* \ StoNED*	– (std.sc)	Grants as input	Grants as output
B&E	– (std.sc)		.094	.092	Engineering	– (std.sc)		.075	.054
	Grants as input	.176		.148		Grants as input	.049		.058
	Grants as output	.186	.293			Grants as output	.179	.207	
Humanities	– (std.sc)		.046	.086	Math. & CS	– (std.sc)		.016	.131
	Grants as input	.116		.115		Grants as input	.074		.136
	Grants as output	.101	.175			Grants as output	.135	.166	
Natural Sc	– (std.sc)		.035	.059	Social Sc	– (std.sc)		.070	.080
	Grants as input	.079		.082		Grants as input	.069		.125
	Grants as output	.135	.188			Grants as output	.122	.161	

*The results in the respective lower triangulars correspond to DEA_{BCC}, whereas the results in the respective upper triangulars correspond to StoNED

Gray-colored cells refer to SDMDs ≤ 0.05

5.3 Results concerning the purpose "estimation of efficiency degrees"

To accurately estimate the efficiency degrees, there are slight differences regarding the results of the methods. For example, the DEA_{BCC}, DEA_{CCR}, and SFA results show higher SDMDs than StoNED and NAA. Therefore, the results for DEA_{BCC} and StoNED are presented as examples in Table 12.

First, we look at the SDMDs for DEA_{BCC}. Some of them are very high, with up to 18.6 percentage points when comparing with the standard scenario. The lowest SDMDs result from the comparison between the standard scenario and the inclusion of grants as input. However, these are not greater than five percentage points for only one variant. Overall, there are clear shifts in the ratios of efficiency degrees to one another. In correspondence to the DEA_{BCC} results in Sect. 5.2, social sciences show the lowest average changes and business & economics the highest. Again, there is no relation with the initial data.

In contrast, as outlined at the beginning, the spread tends to be considerably lower for StoNED. The SDMDs amount to a maximum of 13.1 percentage points when comparing with the standard scenario. Three comparisons yield SDMDs of less than five percentage points. Unlike DEA_{BCC}, engineering and natural sciences have the lowest average SDMDs here. Thus, the research fields with the smallest changes vary with the method used. Overall, the StoNED results differ in the specific values of the SDMDs,

but not in the resulting basic statements. For example, considering grants as input again leads to the smallest changes here.

In summary, for this purpose, bigger changes in the ratios of efficiency degrees result from the consideration of grants as an output rather than as an input. However, even though the inclusion as input yields lower SDMDs, these are still quite high in absolute terms. Although StoNED and NAA show smaller changes than the other methods, the results have no structural differences. This is also valid regarding the research fields. Hence, regarding the considered purpose, both variants of the inclusion of grants can be reasonable.

6 Influence of bibliometric indicators

A series of bibliometric indicators exist that aim to capture the dissemination and reception of research results of academic institutions in the scientific community. Although such indicators have been discussed for some time with regard to their use in evaluations (Clarke 2009; Bornmann et al. 2013; Ketzler and Zimmermann 2013), their wide range is not yet reflected in efficiency analyses (Geissler et al. 2021). Most commonly, the number of publications (cf. the standard scenario) and/or the number of citations[4] (e.g., Bonaccorsi et al. 2006; Wohlrabe and Friedrich 2017) are used. Occasionally, average citations per paper (CPP) are also used (e.g., Agasisti et al. 2012; Mammadov and Aypay 2020). In contrast, more complex bibliometric indicators are rarely applied. For example, the h-index gives the number of publications h with at least h citations (Hirsch 2005) and is thus more robust to outliers. To our knowledge, in efficiency analyses that solely refer to university research, the h-index has so far only been used by Agasisti et al. (2012).

Differences in publication and citation customs between research fields and journals are taken into account by normalized index measures intended to counteract comparability problems.[5] The journal-normalized J-factor (Ball et al. 2009), e.g., relates publications to the respective publication organ. In this way, it considers journal-specific differences in scientific communication. To the best of our knowledge, normalized measures have only been included in efficiency analyses of university research by Geissler et al. (2021). The authors found that the three regarded normalized measures yielded almost identical results in efficiency analyses. Against this background, we focus on considering the J-factor as a representative of normalized index measures. Given the explanations above, we extend the standard scenario

[4] For a critical discussion of the meaning of citations, see, e.g., Fong and Wilhite (2017). It should be noted that, terminologically, citations and the indicators based on them do not represent an output of research but rather present its outcome in the form of an impact indicator. In the context of production-theoretical methods of efficiency measurement, however, this is not important since the classification as an output factor refers to the maximization of this indicator, which is also the case for bibliometric indicators. For a general discussion, see Afsharian et al. (2016) and Dyckhoff (2018).

[5] A list of bibliometric (normalized) measures can be found in, e.g., Van Raan (2005) and Clermont et al. (2017).

Table 13 Considered inputs and outputs for variations of bibliometric indicators

Variants concerning bibliometric indicators	Professors	Research asst	Publications	Citations	CPP	h-index	J-factor
– (std.sc)	I	I	O				
Cit	I	I	O	O			
CPP	I	I	O		O		
h-index	I	I	O			O	
J-factor	I	I	O				O
Cit & CPP	I	I	O	O	O		
Cit & h-index	I	I	O	O		O	
Cit & J-factor	I	I	O	O			O

cit citations

by seven variants of including bibliometric indicators as output factors, as Table 13 shows.

6.1 Descriptive analysis

Based on the selected publication year 2010 and the allocation of journals to research fields as explained in Sect. 3.3, respective citations were gathered by means of the databases of the Competence Centre of Bibliometrics. We collected the citations for the period 2010–2018. For such a period of nine years, Clermont et al. (2021) found sufficient validity for the citation measures considered here, i.e., for CPP, h-index, and J-factor.[6] We calculated these measures for the six research fields and universities.

Table 14 presents the descriptive statistics for the measures. We again observe research field-specific differences, with many citations for engineering and natural sciences, corresponding to their publication numbers. Besides, the average CPP and h-index are the highest there. The J-factors show similar average values due to the normalization across the research fields, but their ranges are quite different, with the largest for humanities.

Table 15 presents the corresponding Pearson correlations. Compared to the factors considered so far, some of the correlations are clearly lower. High correlations of over 70% and even over 90% are found almost exclusively between publications, citations, and the h-index. In contrast, the correlations with CPP and J-factor are considerably lower. In fact, they are often below 40% and not significant. Only in some cases, there are higher correlations. For example, the correlations between CPP and J-factor for mathematics & computer science, natural sciences, and

[6] As a reference group for calculating the J-factor, the citations in the years 2010–2018 for all journal publications were considered which were published by German universities in the same publication year (2010) and the same journal as the respective publication under consideration.

Table 14 Descriptive statistics on the bibliometric indicators*

		Citations	CPP	h-index	J-factor			Citations	CPP	h-index	J-factor
B&E	Min	17	3.8	1	0.3	Engineering	Min	48	5.2	2	0.4
	Max	2,325	72	26	2.5		Max	24,414	84.3	72	1.4
	Mean	462.3	21.4	9.3	1.1		Mean	5,252.5	32.1	29.4	1.0
	Median	306.0	18.0	8	1.1		Median	3,261.0	28.3	28.5	1.0
	Std.dev	486.5	13.4	5.6	0.3		Std.dev	5,514.4	17.4	16.4	0.2
Humanities	Min	0	0	0	0	Math. & CS	Min	46	5.1	3	0.5
	Max	451	36.2	11	12.7		Max	4,717	59.5	39	2.7
	Mean	92.7	6.1	3.7	1.1		Mean	1,003.3	16.4	14.7	1.1
	Median	44.0	4.4	3.0	0.8		Median	664.5	14.2	14.0	1.0
	Std.dev	111.6	6.2	2.6	1.6		Std.dev	909.3	9.4	7.2	0.4
Natural Sc	Min	28	7	3	0.3	Social Sc	Min	0	0	0	0
	Max	67,658	91.4	110	2.2		Max	3,571	67	32	2.1
	Mean	17,902.1	32.7	54.3	1.1		Mean	1,060.9	20.6	15.1	1.1
	Median	15,232.0	32.5	56.0	1.1		Median	825.5	20.6	14.0	1.1
	Std.dev	15,119.3	12.2	25.0	0.2		Std.dev	949.1	10.3	9.0	0.3

*The publication data correspond to 2010 whereas the citations and other bibliometric indicators for these publications are based on 2010–2018

Table 15 Pearson correlations between bibliometric indicators

		Publications	Citations	CPP	h-index	J-factor
B&E	Publications	1				
	Citations	.881	1			
	CPP	*–.035*	.271	1		
	h-index	.886	.899	*.058*	1	
	J-factor	*.064*	.253	.702	*.177*	1
Humanities	Publications	1				
	Citations	.716	1			
	CPP	*.125*	.675	1		
	h-index	.857	.825	.414	1	
	J-factor	*–.049*	*.005*	*.066*	*–.009*	1
Natural Sc	Publications	1				
	Citations	.970	1			
	CPP	.264	.376	1		
	h-index	.959	.943	.344	1	
	J-factor	*.231*	.292	.769	.322	1

		Publications	Citations	CPP	h-index	J-factor
Engineering	Publications	1				
	Citations	.887	1			
	CPP	*.213*	.512	1		
	h-index	.894	.928	.459	1	
	J-factor	.511	.622	.594	.685	1
Math. & CS	Publications	1				
	Citations	.904	1			
	CPP	*.042*	.311	1		
	h-index	.926	.932	*.158*	1	
	J-factor	*.050*	.263	.777	*.161*	1
Social Sc	Publications	1				
	Citations	.939	1			
	CPP	.255	.460	1		
	h-index	.940	.949	.399	1	
	J-factor	*.077*	*.112*	.396	*.129*	1

Most results are significant with $p \leq 0.05$. Non-significant results are written in italics

Gray-colored cells refer to correlations $\geq 90\%$

business & economics are over 70%. When comparing research fields, again engineering and natural sciences show the highest correlations on average.

6.2 Results concerning the purpose "determination of the ranking"

Regarding the purpose of accurately determining the ranking, DEA_{BCC}, DEA_{CCR}, and NAA show very similar results with rather high RCs. In the following, the results for DEA_{CCR} are presented, as this method leads to the highest RCs. When estimating efficiency using SFA in this subsection, infeasibility occurs in many variants. This may be due to a lack of skewness in the initial data (see Ruggiero 2007 for a discussion). Therefore, no reasonable statements can be made for SFA here. Hence, we also present the results for StoNED in the following. The RCs of DEA_{CCR} and StoNED results are presented in Table 16.

For DEA_{CCR}, across all research fields, it can be seen that the RCs are almost always above 70% and often above 90%. Contrastingly, the RCs of the J-factor are comparatively low in the case of two and three outputs. Thus, the J-factor leads to the biggest changes in the rankings overall. For the CPP, the RCs are clearly higher, although these have similarly low correlations as the J-factor in the initial data. If including CPP, h-index, or J-factor in the efficiency analysis, adding citations almost always leads to very high RCs of over 95%. Thus, the ranking hardly

Table 16 Rank correlations for DEA_{CCR} and StoNED when including bibliometric indicators

	DEA$_{CCR}$* \ StoNED*	– (std.sc.)	Cit	CPP	h-index	J-factor	Cit & CPP	Cit & h-index	Cit & J-factor
B&E	– (std.sc.)		.919	.566	.706	.400	.582	.720	.415
	Cit	.905		.791	.772	.516	.812	.854	.601
	CPP	.746	.841		.634	.690	.993	.757	.785
	h-index	.894	.824	.800		.495	.661	.949	.560
	J-factor	.464	.507	.785	.635		.681	.533	.931
	Cit & CPP	.737	.904	.960	.752	.691		.791	.797
	Cit & h-index	.873	.947	.856	.912	.617	.893		.656
	Cit & J-factor	.556	.700	.843	.662	.898	.855	.784	
Humanities	– (std.sc.)		.895	.722	.894	.417	.735	.847	.433
	Cit	.999		.909	.931	.446	.926	.960	.564
	CPP	.997	.996		.836	.384	.991	.872	.524
	h-index	.999	.999	.996		.513	.845	.976	.564
	J-factor	.703	.707	.705	.711		.382	.509	.939
	Cit & CPP	.997	.997	.999	.996	.710		.892	.537
	Cit & h-index	.999	.999	.995	.999	.715	.997		.615
	Cit & J-factor	.703	.708	.705	.711	.999	.711	.716	
Natural Sc	– (std.sc.)		.895	.508	.614	.574	.551	.629	.603
	Cit	.940		.704	.781	.702	.754	.815	.763
	CPP	.993	.941		.758	.786	.982	.808	.840
	h-index	.987	.942	.988		.677	.798	.983	.762
	J-factor	.920	.874	.957	.926		.818	.710	.975
	Cit & CPP	.935	.995	.947	.940	.906		.851	.883
	Cit & h-index	.930	.991	.937	.951	.880	.990		.800
	Cit & J-factor	.870	.930	.911	.884	.952	.960	.933	

	DEA$_{CCR}$* \ StoNED*	– (std.sc.)	Cit	CPP	h-index	J-factor	Cit & CPP	Cit & h-index	Cit & J-factor
Engineering	– (std.sc.)		.943	.720	.709	.353	.745	.696	*.312*
	Cit	.936		.841	.801	.387	.876	.809	.469
	CPP	.902	.900		.882	.600	.985	.886	.712
	h-index	.953	.882	.920		.594	.890	.989	.682
	J-factor	.771	.692	.857	.878		.598	.606	.897
	Cit & CPP	.888	.967	.950	.886	.778		.903	.732
	Cit & h-index	.920	.943	.914	.957	.818	.957		.723
	Cit & J-factor	.779	.778	.860	.874	.957	.865	.892	
Math. & CS	– (std.sc.)		.907	.606	.643	.595	.633	.650	.598
	Cit	.938		.821	.786	.754	.857	.859	.815
	CPP	.896	.957		.743	.839	.988	.850	.897
	h-index	.903	.927	.949		.733	.766	.951	.792
	J-factor	.862	.940	.984	.950		.838	.790	.954
	Cit & CPP	.863	.969	.983	.927	.975		.882	.920
	Cit & h-index	.892	.948	.943	.985	.949	.949		.882
	Cit & J-factor	.847	.950	.973	.932	.990	.987	.953	
Social Sc	– (std.sc.)		.937	.720	.839	.560	.732	.827	.531
	Cit	.977		.869	.878	.623	.879	.914	.648
	CPP	.947	.964		.837	.637	.995	.904	.693
	h-index	.990	.970	.964		.619	.852	.981	.660
	J-factor	.980	.958	.952	.991		.644	.643	.964
	Cit & CPP	.918	.966	.983	.938	.921		.919	.708
	Cit & h-index	.970	.990	.976	.980	.969	.977		.697
	Cit & J-factor	.948	.981	.973	.962	.963	.981	.990	

Almost all results are significant with $p \leq 0.05$. Non-significant results are written in italics

Gray-colored cells refer to correlations $\geq 90\%$

*The results in the respective lower triangulars correspond to DEA_{CCR}, whereas the results in the respective upper triangulars correspond to StoNED

changes. Even when extending the standard scenario only by citations, the RCs are above 90%.

In terms of research fields, the results for social sciences are, on average, the most highly correlated, whereas business & economics have the lowest values. Overall, DEA_{BCC} and NAA widely lead to the same conclusions. For these, the inclusion of the J-factor also induces the biggest changes, whereas including citations results in very high RCs. The highest average RCs partly occur here for other research fields, e.g., concerning DEA_{BCC} for mathematics & computer science.

Regarding StoNED, there are considerably lower RCs of often less than 70%, especially when including the J-factor. However, the inclusion of other bibliometric indicators should also be examined because of the rather low RCs. Only for citations, their additional inclusion leads again to very high RCs of more than 90%. Regarding the research fields with the highest RCs, there are differences compared to DEA_{CCR}. This shows that in our study, there is no structural pattern regarding the research fields.

Table 17 Standard deviations of mean deviation for DEA_{BCC} and NAA when including bibliometric indicators

	DEA_{BCC}* \ NAA*	– (std.sc.)	Cit	CPP	h-index	J-factor	Cit & CPP	Cit & h-index	Cit & J-factor
B&E	– (std.sc.)		.053	.101	.050	.100	.092	.060	.090
	Cit	.123		.084	.052	.086	.059	.029	.060
	CPP	.154	.132		.078	.029	.032	.073	.037
	h-index	.134	.127	.174		.071	.070	.031	.062
	J-factor	.140	.165	.114	.164		.044	.070	.034
	Cit & CPP	.166	.118	.047	.174	.134		.052	.019
	Cit & h-index	.146	.110	.167	.042	.169	.163		.047
	Cit & J-factor	.154	.134	.102	.158	.059	.104	.152	
Humanities	– (std.sc.)		.050	.055	.042	.100	.068	.054	.086
	Cit	.037		.028	.044	.093	.024	.023	.062
	CPP	.044	.029		.031	.082	.020	.016	.055
	h-index	.077	.064	.062		.079	.047	.028	.062
	J-factor	.140	.144	.144	.152		.088	.082	.037
	Cit & CPP	.049	.024	.014	.063	.146		.021	.055
	Cit & h-index	.085	.065	.064	.019	.156	.061		.053
	Cit & J-factor	.143	.140	.140	.147	.032	.140	.147	
Natural Sc	– (std.sc.)		.039	.085	.057	.091	.071	.049	.071
	Cit	.083		.079	.065	.091	.053	.039	.059
	CPP	.082	.046		.044	.028	.030	.047	.028
	h-index	.100	.093	.089		.043	.047	.030	.039
	J-factor	.128	.120	.088	.117		.049	.057	.035
	Cit & CPP	.085	.038	.018	.087	.093		.029	.019
	Cit & h-index	.106	.088	.085	.014	.116	.083		.029
	Cit & J-factor	.126	.113	.084	.114	.019	.086	.112	

	DEA_{BCC}* \ NAA*	– (std.sc.)	Cit	CPP	h-index	J-factor	Cit & CPP	Cit & h-index	Cit & J-factor
Engineering	– (std.sc.)		.039	.051	.027	.050	.051	.035	.043
	Cit	.077		.050	.041	.060	.031	.019	.035
	CPP	.130	.106		.026	.020	.025	.032	.019
	h-index	.073	.061	.104		.026	.033	.025	.024
	J-factor	.185	.207	.172	.177		.041	.043	.030
	Cit & CPP	.131	.105	.010	.103	.174		.017	.014
	Cit & h-index	.078	.058	.103	.010	.179	.102		.017
	Cit & J-factor	.185	.207	.172	.177	0	.174	.179	
Math. & CS	– (std.sc.)		.071	.110	.059	.107	.105	.073	.100
	Cit	.071		.096	.071	.103	.065	.037	.066
	CPP	.094	.057		.076	.039	.043	.076	.047
	h-index	.087	.067	.083		.065	.076	.043	.063
	J-factor	.118	.085	.062	.091		.062	.075	.047
	Cit & CPP	.110	.061	.019	.089	.062		.051	.026
	Cit & h-index	.090	.062	.072	.017	.082	.077		.043
	Cit & J-factor	.120	.086	.062	.093	.004	.062	.083	
Social Sc	– (std.sc.)		.014	.048	.019	.064	.031	.013	.040
	Cit	.056		.063	.027	.071	.033	.015	.045
	CPP	.158	.137		.032	.031	.020	.040	.021
	h-index	.097	.084	.146		.049	.020	.013	.026
	J-factor	.196	.197	.174	.200		.044	.058	.027
	Cit & CPP	.163	.138	.017	.146	.179		.022	.020
	Cit & h-index	.098	.083	.144	.004	.199	.144		.032
	Cit & J-factor	.197	.190	.166	.193	.030	.168	.192	

Gray-colored cells refer to SDMDs ≤ 0.05

*The results in the respective lower triangulars correspond to DEA_{BCC}, whereas the results in the respective upper triangulars correspond to NAA

For the deterministic methods, the inclusion of further bibliometric indicators often leads to rather minor changes in the ranking. An exception is the J-factor, which is why its use as an additional indicator should be examined. For StoNED, considering further bibliometric indicators also seems relevant since the conclusions are similar, but the RCs are lower overall. However, for all methods, considering the frequently used citations does not seem necessary, especially if considering CPP, h-index, or J-factor as well. There are hardly any fundamental differences between the research fields for all the methods.

6.3 Results concerning the purpose "estimation of efficiency degrees"

For this purpose, again similarities in the results between the methods can be observed. For example, DEA_{BCC} and StoNED show comparatively high SDMDs, whereas these are lower for NAA and DEA_{CCR}. In Table 17, the results of DEA_{BCC} and NAA are, therefore, considered as examples.

Once more, we first look at the SDMDs for $DEA_{BCC.}$ With SDMDs of up to 20.7 percentage points, significant shifts in the ratios of the results are evident. Again, the findings are consistent in their basic statements. Thus, the J-factor tends to lead to the highest SDMDs. However, the difference is not as clear as in the case of RCs since the CPP, in particular, also partially lead to high SDMDs. Therefore, the inclusion of all bibliometric indicators should be examined here.

As in the case of RCs, SDMDs are relatively low—mostly below five percentage points—if citations are included as the third output factor when considering CPP, h-index, or J-factor. Meanwhile, considering citations as only additional output factor sometimes leads to SDMDs clearly greater than five percentage points. If we calculate the research field-specific average across all SDMDs, we see that the changes are lowest on average for mathematics & computer science, whereas they are somewhat higher for social sciences and business & economics. This is analogous to the RCs.

For NAA, SDMDs amount up to eleven percentage points, being considerably lower than for DEA_{BCC}. Moreover, they are often lower than five percentage points. The other conclusions made for DEA_{BCC} are also valid for NAA, except that the research fields with the lowest average SDMDs vary with the method used. As can be seen, high SDMDs are observed when the J-factor is used; therefore, for NAA and StoNED, its inclusion should be examined in particular.

A strong dependence on the bibliometric indicators can be stated for an accurate estimation of the efficiency degrees, whereby the J-factor again leads to the biggest changes here. As in the case of RCs, the inclusion of citations seems to be negligible, at least when considering CPP, h-index, or J-factor as well. The basic results here also suggest a dependence of the results on the considered research field and the method used. However, apart from the basic differences in height, no pattern is discernible, e.g., between the research fields.

7 Discussion and conclusions

We investigated the impact of the inclusion of selected input and output factors based on a standard scenario depending on the purpose of the analysis, the research field under consideration, and the method used. We addressed the effects of taking into account different types of expenditures, grants, and bibliometric indicators that have hardly been included in efficiency analyses so far. In the following, we will discuss which fundamental implications and guidance can be derived from our results for evaluators as well as university and political decision-makers.

To avoid misinterpretation of our results, it must be clarified that we do not know the true efficiency of the research fields and universities under consideration. Therefore, we cannot provide any information about which input and output factors or methods have to be selected to model the actual efficiency as accurately as possible. Our conceptually oriented analyses rather answer the question of to what extent various extensions of the standard scenario lead to a change in the results of the efficiency measurement. As mentioned in Sect. 5, the reasonability of including input and output factors not only depends on the RCs and SDMDs but especially on the

goals pursued in the respective situation. Generally, an analysis should always follow the pursued goals whereas the used input and output factors are—more or less suitable—indicators to quantify these goals (Ahn and Clermont 2018). In this context, the collection of data for input and output factors usually requires a high effort that is only reasonable when the additional input and output factors lead to substantial changes in the results. Our analyses show that this is not always the case.

If the addition of a certain input or output factor leads to small changes, represented by high RCs or low SDMDs, the results show strong similarities with respect to the purpose considered. This indicates that the additional inclusion of the corresponding factor is not necessary from a result-driven point of view. The reverse is true in the case of strong changes between the results of the considered variants, i.e., low RCs and high SDMDs. In this case, the additional inclusion of the input or output factor has a stronger effect on the results, i.e., the inclusion of such factors should be examined.

The question arises whether there is a relation between the correlation in the initial data and the similarity of the results of the respective variants considered. Concerning DEA, Dyson et al. (2001) stated that such a conclusion is not possible due to the lack of translation invariance for additive shifts. Our investigations confirm this finding for the other methods considered here, independent of the underlying purpose of the analysis and the factors included. For example, the initial data showed comparatively low correlations between equipment expenditures and other expenditure types. However, this did not mean that considering equipment expenditures would have led to notably lower RCs. Furthermore, for all methods, there are comparisons in which one factor is less correlated than another in the initial data, but leads to higher RCs or lower SDMDs in the results.

Our analyses naturally yield different evaluation criteria values in detail, depending on the research field and method considered and the variants compared. Moreover, there are also differences in the average level of values. However, it is not found that one research field or method leads to higher or lower RCs/SDMDs than the others for all investigations. Regarding the input and output factors considered, largely consistent tendencies are identified, i.e., the resulting core statements regarding the compared variants apply to all research fields and methods. This applies, for example, to the question of which input or output factor leads to the greatest changes. Consequently, statements can be derived below that are valid regardless of the considered research field and selected method for measuring efficiency.

The resulting key findings of our investigations are summarized in Table 18, which can serve as guidance for efficiency analyses of research at universities. The table provides an overview of the input and output factors for which an examination of their use in efficiency analyses seems particularly expedient since their consideration leads to big changes in the results. It is differentiated according to the two purposes under consideration, as there are differences in the results and thus also in the implications. The results in the second column correspond to the purpose of accurately determining the ranking (Sects. 4.2, 5.2, and 6.2) while the results in the third column refer to the accurate estimation of efficiency degrees (Sects. 4.3, 5.3, and 6.3). A further differentiation is made between the methods whenever the average heights of the RCs and SDMDs varied strongly between them.

Table 18 Central implications of the investigations

	Considered purpose of an efficiency measurement	
	Accurate determination of the ranking	Accurate estimation of efficiency degrees
Expenditures	They seem to be neglectable	The inclusion should be examined, especially regarding equipment expenditures combined with DEA_{BCC} and DEA_{CCR}
Grants	The inclusion as input factor seems to be neglectable The inclusion as output factor should be examined	The inclusion as input factor and as output factor should be examined
Bibliometric indicators	The inclusion of citations seems to be neglectable For StoNED, the inclusion of the other indicators should be examined For DEA_{BCC}, DEA_{CCR}, and NAA, the inclusion of the J-factor should be examined	For DEA_{BCC} and StoNED, the inclusion of all regarded indicators should be examined For DEA_{CCR} and NAA, the inclusion of the J-factor should be examined When already including CPP, h-index, or J-factor, citations seem to be neglectable

In summary, there are differences regarding the purposes pursued by the efficiency measurement, especially in their sensitivities to additional input/outputs factors, which illustrates the relevance for selecting factors. Thus, examining the inclusion of additional input and output factors appears to be particularly appropriate for accurately estimating the efficiency degrees. For this purpose, it can be stated that the focus should be on considering equipment expenditures, grants (both as input and output), and the J-factor. Our results indicate that the inclusion of grants as an output and the J-factor should be examined mainly for the purpose of accurately determining the ranking.

8 Limitations and outlook

The findings of our study are limited to the focus set. In this respect, various extensions are possible to develop additional guidance for efficiency analyses of university research. For example, an extension to other research fields, methods, and input or output factors is obvious. Time-series analyses could also be of interest to examine the robustness of our findings over time. Moreover, considering further purposes and evaluation criteria seems fruitful. For example, it should be noted that the two evaluation criteria we use—RC and SDMD—only provide information regarding average similarity. Nevertheless, individual stronger deviations may occur.

In general, there are no overall differences in the methods' level of RCs and SDMDs. In detail, however, it is visible that the results differ strongly across the methods. While this was to be expected in accordance with respective literature (e.g., Ahn et al. 2020), it does not provide information on which method

performs better in estimating efficiency. This emphasizes the importance of the method choice for the present situation, which is a separate research topic. As it is addressed in only a few comprehensive studies that use Monte Carlo simulations to evaluate a method's performance in different scenarios (e.g., Andor and Hesse 2014), this seems to be a fruitful path for future research.

Regarding the bibliometric indicators considered, various further investigations can be carried out. For example, instead of the journal-normalized J-factor, the inclusion of the field-normalized Mean Normalized Citation Score (Waltman et al. 2011) could be analyzed. In contrast to journal normalization, field normalization allocates publications based on defined categories (e.g., university departments). One aspect that has been often discussed in this regard is the allocation of publications to research fields. Ultimately, this can only be ensured through the standardization of data collection, which is currently not foreseeable. Furthermore, questions arise about the relationship between input factors and bibliometric indicators. It could be investigated, e.g., to what extent an increase in staff of institutes increases their attractiveness for professors who are particularly strong in research. In addition to an expected increase in publications, this could also increase CPP, h-index, and J-factor.

Altmetrics represent a relatively new category of indicators for measuring the perception of research performance, especially by the interested public. Here, e.g., tweets, news articles, blog posts, and Mendeley readership are counted. Analyzing the inclusion of such indicators could also lead to interesting insights. Additionally, output factors that reflect the third mission could be considered. One such aspect is the transfer of scientific achievements into practice, measured, for example, by the number of patents, licenses, and start-ups.

Funding Open Access funding enabled and organized by Projekt DEAL. This work was supported by the German Federal Ministry of Education and Research under Grant number 16PU17001. The responsibility of the content lies with the authors.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

Abramo G, D'Angelo CA (2009) Assessing technical and cost efficiency of research activities: A case study of the Italian university system. Research Evaluation 18:61–70. https://doi.org/10.3152/095820209X408869

Afsharian M, Ahn H, Neumann L (2016) Generalized DEA: An approach for supporting input/output factor determination in DEA. Benchmarking 23:1892–1909. https://doi.org/10.1108/BIJ-07-2015-0074

Agasisti T, Catalano G, Landoni P, Verganti R (2012) Evaluating the performance of academic departments: An analysis of research-related output efficiency. Research Evaluation 21:2–14. https://doi.org/10.1093/reseval/rvr001

Ahn H, Clermont M (2018) Performance Measurement. In: Corsten H, Gössinger R, Spengler TS (eds) Handbuch Produktions- und Logistikmanagement in Wertschöpfungsnetzwerken. De Gruyter Oldenbourg, Berlin, Boston, pp 886–903

Ahn H, Le MH (2015) DEA efficiency of German savings banks: Evidence from a goal-oriented perspective. J Bus Econ 85:953–975

Ahn H, Clermont M, Langner J (2020) Verfahren zur Effizienzanalyse von Hochschulen: Überblick und empirischer Vergleich. Qualität in Der Wissenschaft 14:112–119

Ahn H, Dyckhoff H, Gilles R (2007) Datenaggregation zur Leistungsbeurteilung durch Ranking: Vergleich der CHE- und DEA-Methodik sowie Ableitung eines Kompromissansatzes. Z Betriebswirt 77:615–643. https://doi.org/10.1007/s11573-007-0046-z

Ahn H, Höfer-Diehl Y, Neumann L, Vazquez Novoa N (2012) Fakultätsübergreifende Beurteilung der Lehrleistung: Eine Methodik zum Vergleich der Lehreffektivität und Lehreffizienz von Studiengängen. Hochschulmanagement 7:110–115

Aigner D, Lovell CAK, Schmidt P (1977) Formulation and estimation of stochastic frontier production function models. Journal of Econometrics 6:21–37. https://doi.org/10.1016/0304-4076(77)90052-5

Andor M, Hesse F (2014) The StoNED age: The departure into a new era of efficiency analysis? A Monte Carlo comparison of StoNED and the "oldies" (SFA and DEA). J Prod Anal 41:85–109. https://doi.org/10.1007/s11123-013-0354-y

Archambault É, Beauchesne OH, Caruso J (2011) Towards a multilingual, comprehensive and open scientific journal ontology. In: Noyons B, Ngulube P, Leta J (eds) Proceedings of the 13th International Conference of the International Society for Scientometrics and Informetrics, pp 66–77

Ball R, Mittermaier B, Tunger D (2009) Creation of journal-based publication profiles of scientific institutions: A methodology for the interdisciplinary comparison of scientific research based on the J-factor. Scientometrics 81:381–392. https://doi.org/10.1007/s11192-009-2120-5

Banker RD, Charnes A, Cooper WW (1984) Some models for estimating technical and scale inefficiencies in data envelopment analysis. Manage Sci 30:1078–1092. https://doi.org/10.1287/mnsc.30.9.1078

Battese GE, Coelli TJ (1988) Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. Journal of Econometrics 38:387–399. https://doi.org/10.1016/0304-4076(88)90053-X

Beasley JE (1990) Comparing university departments. Omega 18:171–183. https://doi.org/10.1016/0305-0483(90)90064-G

Berbegal-Mirabent J, Solé Parellada F (2012) What are we measuring when evaluating universities' efficiency? Reg Sect Econ Stud 12:31–46

Bielecki A, Albers S (2012) Eine Analyse der Forschungseffizienz deutscher betriebswirtschaftlicher Fachbereiche basierend auf den Daten des Centrums für Hochschulentwicklung (CHE). Christian-Albrechts-Universität zu Kiel, Arbeitspapiere des Instituts für Betriebswirtschaftslehre

Bogetoft P, Otto L (2011) Benchmarking with DEA, SFA, and R. International series in operations research & management science, vol 157. Springer, New York

Bonaccorsi A, Daraio C, Simar L (2006) Advanced indicators of productivity of universities: An application of robust nonparametric methods to Italian data. Scientometrics 66:389–410. https://doi.org/10.1007/s11192-006-0028-x

Bornmann L, Mutz R, Daniel H-D (2013) Multilevel-statistical reformulation of citation-based university rankings: The Leiden ranking 2011/2012. J Am Soc Inf Sci Tec 64:1649–1658. https://doi.org/10.1002/asi.22857

Büschken J (2009) When does data envelopment analysis outperform a naïve efficiency measurement model? Eur J Oper Res 192:647–657. https://doi.org/10.1016/j.ejor.2007.09.026
Charnes A, Cooper WW, Rhodes E (1978) Measuring the efficiency of decision making units. Eur J Oper Res 2:429–444. https://doi.org/10.1016/0377-2217(78)90138-8
Clarke R (2009) A citation analysis of Australian information systems researchers: Towards a new ERA? Australas J Inf Syst 15:23–44. https://doi.org/10.3127/ajis.v15i2.465
Clermont M (2016) Effectiveness and efficiency of research in Germany over time: An analysis of German business schools between 2001 and 2009. Scientometrics 108:1347–1381. https://doi.org/10.1007/s11192-016-2013-3
Clermont M, Dirksen A, Scheidt B, Tunger D (2017) Citation metrics as an additional indicator for evaluating research performance? An analysis of their correlations and validity. Bus Res 10:249–279. https://doi.org/10.1007/s40685-017-0044-0
Clermont M, Krolak J, Tunger D (2021) Does the citation period have any effect on the informative value of selected citation indicators in research evaluations? Scientometrics 126:1019–1047. https://doi.org/10.1007/s11192-020-03782-1
Cook WD, Green RH, Zhu J (2006) Dual-role factors in data envelopment analysis. IIE Trans 38:105–115. https://doi.org/10.1080/07408170500245570
De Witte K, López-Torres L (2017) Efficiency in education: A review of literature and a way forward. Journal of the Operational Research Society 68:339–363
De Fraja G, Facchini G, Gathergood J (2019) Academic salaries and public evaluation of university research: Evidence from the UK research excellence framework. Econ Policy 34:523–583. https://doi.org/10.1093/epolic/eiz009
De la Torre EM, Agasisti T, Perez-Esparrells C (2017) The relevance of knowledge transfer for universities' efficiency scores: An empirical approximation on the Spanish public higher education system. Research Evaluation 26:211–229. https://doi.org/10.1093/reseval/rvx022
Duan SX (2019) Measuring university efficiency: An application of data envelopment analysis and strategic group analysis to Australian universities. Benchmarking 26:1161–1173. https://doi.org/10.1108/BIJ-10-2017-0274
Dyckhoff H (2018) Multi-criteria production theory: Foundation of non-financial and sustainability performance evaluation. J Bus Econ 88:851–882. https://doi.org/10.1007/s11573-017-0885-1
Dyson RG, Allen R, Camanho AS, Podinovski VV, Sarrico CS, Shale EA (2001) Pitfalls and protocols in DEA. Eur J Oper Res 132:245–259. https://doi.org/10.1016/S0377-2217(00)00149-1
Expósito-García A, Velasco-Morente F (2018) How efficient are universities at publishing research? A data envelopment analysis of Spanish state universities. El Profesional De La Información 27:1108–1115. https://doi.org/10.3145/epi.2018.sep.14
Fong EA, Wilhite AW (2017) Authorship and citation manipulation in academic research. PLoS ONE 12:e0187394. https://doi.org/10.1371/journal.pone.0187394
Furková A (2013) Alternative approaches to efficiency evaluation of higher education institutions. ERIES Journal 6:167–178. https://doi.org/10.7160/eriesj.2013.060304
Geissler M, Gralka S, Wohlgemuth M (2021) Effizienzmessung im Hochschulsektor: Ein Vergleich zitationsbasierter Output-Indikatoren. Hochschulmanagement 16:14–23
Ghimire S, Amin SH, Wardley LJ (2021) Developing new data envelopment analysis models to evaluate the efficiency in Ontario Universities. J Informet. https://doi.org/10.1016/j.joi.2021.101172
Gralka S, Wohlrabe K, Bornmann L (2019) How to measure research efficiency in higher education? Research grants vs. publication output. J High Educ Policy Manag 41:322–341. https://doi.org/10.1080/1360080X.2019.1588492
Gralka S (2018) Stochastic frontier analysis in higher education: A systematic review. CEPIE Working Papers, Technische Universität Dresden
Henningsen A, Bělín M, Henningsen G (2017) New insights into the stochastic ray production frontier. Econ Lett 156:18–21. https://doi.org/10.1016/j.econlet.2017.04.006
Hirsch JE (2005) An index to quantify an individual's scientific research output. Proc Natl Acad Sci U S A 102:16569–16572. https://doi.org/10.1073/pnas.0507655102
Ibrahim C, Fadhli R (2021) Performance of Indonesia's world-class university efficiency with bibliometrics (scientific strength) approach and data envelopment analysis. WEB 18:32–50. https://doi.org/10.14704/WEB/V18I1/WEB18003
Jauhar SK, Pant M, Dutt R (2018) Performance measurement of an Indian higher education institute: A sustainable educational supply chain management perspective. International Journal of System Assurance Engineering and Management 9:180–193. https://doi.org/10.1007/s13198-016-0505-4

Jiang J, Lee SK, Rah M-J (2020) Assessing the research efficiency of Chinese higher education institutions by data envelopment analysis. Asia Pacific Educ Rev 21:423–440. https://doi.org/10.1007/s12564-020-09634-0
Johnes G (1995) Scale and technical efficiency in the production of economic research. Appl Econ Lett 2:7–11. https://doi.org/10.1080/135048595357717
Johnes G (1997) Costs and industrial structure in contemporary British higher education. Econ J 107:727–737
Johnes J (2006) Data envelopment analysis and its application to the measurement of efficiency in higher education. Econ Educ Rev 25:273–288. https://doi.org/10.1016/j.econedurev.2005.02.005
Johnes G, Johnes J (1992) Apples and oranges: The aggregation problem in publications analysis. Scientometrics 25:353–365. https://doi.org/10.1007/BF02028091
Johnes G, Johnes J (1993) Measuring the research performance of UK economics departments: An application of data envelopment analysis. Oxf Econ Pap 45:332–347. https://doi.org/10.1093/oxfordjournals.oep.a042095
Johnes J, Johnes G (1995) Research funding and performance in U.K. university departments of economics: A frontier analysis. Econ Educ Rev 14:301–314. https://doi.org/10.1016/0272-7757(95)00008-8
Johnes J, Yu L (2008) Measuring the research performance of Chinese higher education institutions using data envelopment analysis. China Econ Rev 19:679–696. https://doi.org/10.1016/j.chieco.2008.08.004
Jondrow J, Lovell CAK, Materov IS, Schmidt P (1982) On the estimation of technical inefficiency in the stochastic frontier production function model. Journal of Econometrics 19:233–238. https://doi.org/10.1016/0304-4076(82)90004-5
Ketzler R, Zimmermann KF (2013) A citation-analysis of economic research institutes. Scientometrics 95:1095–1112. https://doi.org/10.1007/s11192-012-0850-2
Kounetas K, Anastasiou A, Mitropoulos P, Mitropoulos I (2011) Departmental efficiency dif-ferences within a Greek university: An application of a DEA and Tobit analysis. Int Trans Oper Res 18:545–559. https://doi.org/10.1111/j.1475-3995.2011.00813.x
Kuosmanen T, Kortelainen M (2012) Stochastic non-smooth envelopment of data: Semi-parametric frontier estimation subject to shape constraints. J Prod Anal 38:11–28. https://doi.org/10.1007/s11123-010-0201-3
Lee BL, Worthington AC (2016) A network DEA quantity and quality-orientated production model: An application to Australian university research services. Omega 60:26–33. https://doi.org/10.1016/j.omega.2015.05.014
Linton JD, Tierney R, Walsh ST (2011) Publish or perish: How are research and reputation related? Ser Rev 37:244–257. https://doi.org/10.1016/j.serrev.2011.09.001
Löthgren M (1997) Generalized stochastic frontier production models. Econ Lett 57:255–259. https://doi.org/10.1016/S0165-1765(97)00246-2
Mammadov R, Aypay A (2020) Efficiency analysis of research universities in Turkey. Int J Educ Dev 75:1–11. https://doi.org/10.1016/j.ijedudev.2020.102176
McMillan ML, Datta D (1998) The relative efficiencies of Canadian universities: A DEA perspective. Canadian Public Policy/analyse De Politiques 24:485–511. https://doi.org/10.2307/3552021
Meeusen W, Van den Broeck J (1977) Efficiency estimation from Cobb-Douglas production functions with composed error. Int Econ Rev 18:435–444. https://doi.org/10.2307/2525757
Olson JA, Schmidt P, Waldman DM (1980) A Monte Carlo study of estimators of stochastic frontier production functions. Journal of Econometrics 13:67–82. https://doi.org/10.1016/0304-4076(80)90043-3
Rayeni MM, Saljooghi FH (2010) Network data envelopment analysis model for estimating effi-ciency and productivity in universities. J Comput Sci 6:1252–1257. https://doi.org/10.3844/jcssp.2010.1252.1257
Ruggiero J (2007) A comparison of DEA and the stochastic frontier model using panel data. Int Trans Oper Res 14:259–266. https://doi.org/10.1111/j.1475-3995.2007.00585.x
Schaefer J, Clermont M (2018) Stochastic non-smooth envelopment of data for multi-dimensional output. J Prod Anal 50:139–154. https://doi.org/10.1007/s11123-018-0539-5
Sexton TR, Silkman RH, Hogan AJ (1986) Data envelopment analysis: Critique and extensions. New Directions for Program Evaluation:73–105. https://doi.org/10.1002/ev.1441
Spearman C (1904) The proof and measurement of association between two things. Am J Psychol 15:72–101. https://doi.org/10.2307/1412159

Tunger D, Ahn H, Clermont M, Krolak J, Langner J, Mittermaier B (2022) Altmetrics as additional indicators for analyzing research efficiency of universities? A conceptual and empirical study. Accepted by: Betriebswirtschaftliche Forschung und Praxis

Van Raan AFJ (2005) Measuring science. In: Moed HF, Glänzel W, Schmoch U (eds) Handbook of quantitative science and technology research: The use of publication and patent statistics in studies of S&T systems, vol 415. Kluwer Academic Publishers, Dordrecht, pp 19–50

Waltman L, van Eck NJ, van Leeuwen TN, Visser MS, van Raan AF (2011) Towards a new crown indicator: Some theoretical considerations. J Informet 5:37–47. https://doi.org/10.1016/j.joi.2010.08.001

Wang DD (2019) Performance-based resource allocation for higher education institutions in China. Socioecon Plann Sci 65:66–75. https://doi.org/10.1016/j.seps.2018.01.004

Wang X, Hu H (2017) Sustainability in Chinese higher educational institutions' social Science research: A performance interface toward efficiency. Sustainability. https://doi.org/10.3390/su9111952

Wohlrabe K, Friedrich E (2017) The efficiency of economics departments reconsidered. Economics Bulletin 37:1602–1611

Zhang L, Bao W, Sun L (2016) Resources and research production in higher education: A longitudinal analysis of Chinese universities, 2000–2010. Res High Educ 57:869–891