Business Research (2017) 10:249–279
DOI 10.1007/s40685-017-0044-0

ORIGINAL RESEARCH

Citation metrics as an additional indicator for evaluating research performance? An analysis of their correlations and validity

Marcel Clermont[1] · **Alexander Dirksen**[2] · **Barbara Scheidt**[3] · **Dirk Tunger**[3]

Received: 11 August 2015 / Accepted: 20 January 2017 / Published online: 10 February 2017

Abstract This paper investigates the potential expansion of an indicator set for research performance evaluation to include citations for the mapping of research impact. To this end, we use research performance data of German business schools and consider the linear correlations and the rank correlations between publication-based, supportive, and citation-based indicators. Furthermore, we compare the business schools in partial ratings of the relative indicators amongst themselves and with those business schools that are classified in other studies as being strong in research and/or reputable. Only low correlations are found between the citation metrics and the other indicator types. Since citations map research outcome, this is an expected result in terms of divergent validity. Amongst themselves, the citation metrics display high correlations, which, in accordance with the convergent validity, shows that they can represent research outcome. However, this does not apply to the J-factor, which is a journal-based normalizing citation metric.

Keywords Citation metrics · German business schools · Research performance evaluation · Research ranking · Validity

JEL Classification H83 · I23

✉ Marcel Clermont
m.clermont@tu-braunschweig.de

[1] Institute of Management Control and Business Accounting, Technische Universität Braunschweig, Fallersleber-Tor-Wall 23, 38100 Brunswick, Germany

[2] Sparkassenverband, Westfalen-Lippe, Regina-Protmann-Straße 1, 48159 Münster, Germany

[3] Forschungszentrum Jülich GmbH, Central Library, Bibliometrics Team, 52425 Jülich, Germany

1 Introduction

In recent years, various socioeconomic developments, especially in the political sphere, and not least an increasing internationalization and harmonization of university performance, have led to numerous changes in the European university sector (De Filippo et al. 2012; Delgado-Márquez et al. 2013). In this context, Iñiguez De Onzoño and Carmona (2007), for example, identify a clear trend towards the model of the competitive American academic market with European universities increasingly adopting American practices and attaching great value to quantifiable research performance. In Germany, this is manifested, for instance, in new competitive structures and an increased implementation of market-based instruments and mechanisms (Winkler 2014). This is also noted by the Federal Ministry of Education and Research (2010) in its call for applications on the topic of economics of science: "It can consequently be assumed that with respect to their operations and structures in the fields of research and teaching the universities have considerable potential for improvement." Various ratings and rankings represent a first starting point for the federal and state governments to obtain information on the research performance of their universities.

A frequently criticized aspect is the choice of criteria for evaluating research performance. In this context, various rankings weight publication indicators relatively highly, but ignore the impact of publication performance. An indicator frequently discussed in this context by scientometric researchers is the number of citations that an article receives (Van Raan 1996). Nosek et al. (2010) draw attention to the fact that citations represent an impact indicator which is "valid, relatively objective, and, with existing databases and search tools, straightforward to compute".

Since various rankings do not consider citation metrics, the question arises of whether the survey and evaluation of citation metrics lead to a meaningful, in the sense of substantial and desirable, extension of the research rankings. However, it is not easy to answer this fundamental question. This is mainly due to the fact that the "real" research performance of the university or faculty is unknown. Attempts are made to draw conclusions about research performance by employing surrogates in the form of measurable indicators. Even if, in formal terms, a citation is only a reference to a publication in the bibliography of another publication, the citation nevertheless symbolizes that a flow of information or a perception and utilization of the information and/or research findings has taken place (Stock 2001). Accordingly, the utilization of citations, especially as part of a comprehensive set of indicators, generates in principle a more complete, more characteristic picture of the research profile of a university or faculty (Jensen et al. 2009; Clermont and Dirksen 2016). With respect to the conception of the corresponding impact indicators, different citation metrics have been proposed and discussed in the scientometric literature. However, including several citation metrics in a research ranking does not appear meaningful from our perspective since the aspect of the impact of research performance in a ranking would then be weighted disproportionately heavily.

To obtain some evidence for the value of an additional use of citation metrics in research rankings, we take a look at the research ranking of German business

schools (BuSs) conducted by the Centre for Higher Education (CHE). As yet, the CHE has not used any citation metrics to analyse the research performance of BuSs. Apart from this neglect of citation metrics, the CHE displays a multidimensional structure of indicators for representing research performance. The CHE ranking thus enables differentiated statements to be made about the performance structure and achievement of BuSs and this will, therefore, be the focus of the following analyses. The data required for the evaluation are collected at three-year intervals, evaluated, and published in a popular form.

Consideration is given to the construct validity to analyse the question of the extent to which citation metrics represent a meaningful supplement to the CHE research ranking of BuSs. In this context, we examine the relation between the different citation metrics and the research performance indicators currently employed by the CHE. We, therefore, intend to answer the following questions in our study:

> What is the relation of the CHE research performance indicators to potential citation metrics? Do citation metrics represent valid indicators for mapping research performance?

To answer these research questions, our paper is structured as follows: In the next section, we will give an overview of the present state of the art, firstly with regard to the discussion of research indicators and secondly with regard to studies analysing the relation between different citation metrics. In this context, we also discuss the aspect of the validity of research performance indicators. In Sect. 3, we present our study design. The results of our analysis will be shown and discussed in Sects. 4 and 5. Our paper concludes with a discussion of implications and limitations as well as an outlook for further (potential) research issues.

2 Indicator-based performance measurement at universities

2.1 Objective and indicator system of business administration research performance

The identification, evaluation, and control of performance require the establishment of standards which must ultimately be based on the fundamental objectives of the relevant policymakers and target groups if they are to be accepted and achieve the desired incentive effect (Keeney and Raiffa 1993). The research objectives formulated generally in university statutes are fundamental with respect to the interests of various stakeholders; however, they are not specific and manageable enough for performance measurement. Hardly any studies are to be found on the systematic, explicit derivation of specific fundamental objectives of performance in the university sector (Ahn et al. 2012).

According to Chalmers (1990), the basic objective of the academic research consists of the generation, publication, and exploitation of new knowledge about the world. The success of such an objective can only be measured indirectly, which is why different metrics and indicators are proposed and discussed in the literature. To

systematize such indicators, Lorenz and Löffler (2015) distinguish between productivity, impact, and esteem indicators. Under the heading of productivity indicators, they subsume those indicators which map the publication output of an academic organizational unit (e.g. number of published journal articles). Impact indicators, in contrast, provide information on the perception of publication output in the scientific community (e.g. number of citations). The authors summarize surrogates for the quality of research work under the heading of esteem indicators, such as honorary doctorates and memberships of editorial boards of journals. Dyckhoff et al. (2005a) provide a more extensive structuring of research indicators, the content of which, moreover, is based on the objectives of research performance in business administration. In particular, they distinguish between main effects and desirable and undesirable side effects as the outputs of academic research and expenditure factors as input. In contrast to Lorenz and Löffler, they do not only focus on publication- or citation-based indicators, but also include other research factors such as acquisition of third-party funding and education of young scientists.

Due to the large number of possible indicators for measuring performance in the university sector, as part of a Delphi study, Palomares-Montero and García-Aracil (2011) asked various university stakeholders in Spain what indicators they considered to be key performance indicators. With respect to research performance, the number of publications and citations proved to be the central indicators. However, the number of PhDs and the volume of third-party funding were also regarded as relevant.

The significance of publication-based indicators derived at the university level can hardly be disputed even for business administration. The distribution and discussion of new knowledge in business administration are mainly undertaken by means of written papers. Although great significance is attached to publication in scientific journals (e.g. Albers 2015; Bort and Schiller-Merkens 2010), the lack of interest in other publication options, such as monographs or contributions to collections, is also criticized (e.g. Dilger and Müller 2012). Furthermore, publications are generally classified or weighted according to certain criteria with respect to their value. However, such weighting of written research work is a matter of controversy (for a related overview, see Clermont and Dirksen 2016 and the literature cited therein). Since, for example, the journal rating JOURQUAL2 is used to weight the quality of journal articles, Eisend (2011) investigated the validity of this rating on the basis of a correlation study with other rankings and ratings already established in the science community, whereas Lorenz and Löffler (2015) analysed the robustness of the methodology of the Handelsblatt rankings.

The use of citations as indicators of research performance is based on the assumption that the frequency of citations demonstrates that the progress in knowledge contributed by a paper is imparted to other scientists by their study of this paper. Conversely, uncited papers mean that they do not provide a major contribution for other scientists (see, e.g. Schmitz 2008 and the literature cited therein).[1] The degree of significance attributed to citations is, however, also subject to criticism, which is why some authors reject the use of citation metrics as a matter

[1] For an extensive discussion of the use of citation-based indicators for measuring the research performance of business administration scientists (in Germany), see Schmitz (2008) or Müller (2012).

of principle (e.g. MacRoberts and MacRoberts 1996). Thus, for example, due to the great effort involved in an evaluation of content, negative citations are usually not eliminated from the analyses (see Weinstock 1971, for an overview of different citation reasons). In addition, there are also opportunities for manipulation, for example, by self-citations. Although this can be combatted by a rigorous exclusion of all self-citations (Dilger 2010), this also removes justified self-citations (Dilger 2000).[2] To avoid such problems and to counteract criticism, especially in bibliometrics, various indicators have been proposed, some of which we will discuss in more detail in Sect. 3.

An indicator of research performance frequently used in German science policy is the number of PhD dissertations. By definition, PhD dissertations serve the education and further training of young scientists. The aim of a PhD student is to generate new scientific knowledge and to present this in the form of a dissertation for publication. The number of successful PhD candidates may, therefore, indicate that a BuS is achieving its research objective. Third-party funds are regarded as predicting desirable research performance. There is a positive attitude on the part of science policy in general and university management in particular to the procurement of third-party funds, which is why it appears justifiable to include such funds as a desirable performance indicator.

It should be noted that relations between the above-mentioned indicators are conceivable. For example, past research achievements and, in particular, publications may lead to more research projects financed by third-party funds, since past research performance can be regarded by providers of third-party funding as a predictor of future research achievements. These funds are used to employ new staff, which in turn increases the number of PhD dissertations. However, some providers of third-party funding, such as the German Research Foundation (DFG), are aware of this problem, which is why applicants are only allowed to specify the five most important publications when applying for project funds (Kleiner 2010). Furthermore, the argument for the acquisition of third-party funds from industry will probably carry less weight. As a rule, the publication of research findings tends to play a minor role in industry. In this case, factors such as density of companies and universities in the region and previous experience with project work would probably be more decisive.

2.2 Validity of citation metrics[3]

Both construction- and application-related conditions must be fulfilled to ensure the quality of an indicator-based representation of a state of affairs that cannot be

[2] In general, in bibliometric analyses of citations the decisive aspect is not to consider individual publications or authors but to concentrate on a statistically relevant parent population. Although the above-mentioned problems are still associated with such a set nevertheless they do not falsify the informative value or only to a lesser extent. Hence, bibliometrics is to be regarded as a method of making data-based and therefore objectified statements, supported by a large data basis, on the perception and impact of scientific publications.

[3] We would like to thank an anonymous reviewer for the idea of integrating the discussion about validity into our manuscript.

directly observed as part of a performance analysis (Rassenhövel 2010). The central construction-related condition is the validity of an indicator. It is a measure of its descriptive quality, that is to say the agreement of representation and reality. An indicator is valid if it actually reflects the state of affairs designated by the defined concept (Kromrey and Roose 2016). In the context of performance surveys, Messick (1995, p. 5) notes that validity is in fact a social value that has "meaning and force whenever evaluative judgments and decisions are made." Numerous types of validity are used in the literature and their meaning depends on the respective context.

In the present context, the construct validity is of major relevance (see Cronbach and Meehl 1955).[4] In verifying the construct validity, we differentiate between convergent and discriminant validity in the following. Convergent validity indicates the degree to which different measures of the same construct correlate with each other. Conversely, discriminant validity means the degree to which measures of different constructs differ from each other (Campbell and Fiske 1959).

If the construct validity is transferred to the case of an integration of citation metrics into CHE's existing indicator system under consideration here, then it is not clear whether the citation metrics should be convergently or divergently valid with respect to the indicators previously used. The assignment depends on whether citations can be regarded as representing research performance or whether they measure something else which is independent of the content or meaning of research performance. This is ultimately decided by how performance is defined. According to Neely et al. (1995), the criteria for evaluating performance are effectiveness and efficiency. Both terms largely refer to the production process and thus above all to input, throughput, and output (for a general definition, see, e.g. Ahn and Dyckhoff 2004). Citations, however, are related to the outcome, i.e. the effect of publications, as a component of performance.

Furthermore, the productive unit under consideration should also be able to influence the components of performance. This is basically the case with indicators such as number of publications or number of PhDs, because a good result can be achieved by hard work. In contrast, apart from self-citations, citations of an article cannot be influenced by the author or the faculty. They depend, among other things, on the specific research field. If, for example, an author works and publishes in a strongly focused research field, then there are naturally a greater number of other authors who can perceive his work and cite it. Additionally, the number of citations depends on the type of paper; papers discussing the state of the art of a research field are on average cited more frequently than original research papers. For example, if all the articles listed in the Science Citation Index Expanded of the Web of Science (WoS) from the year 2010 are taken into consideration, then the average citation rate of the 63,582 reviews is almost three times as high as the 1,186,181 articles. In this sense, Webster et al. (2009) showed that in the research field of evolutionary psychology those articles that had an extensive list of references were cited significantly more frequently.

[4] For an application and examination of criteria of construct validity to the context of journal rankings, see Eisend (2011).

For these reasons, we believe that citations do not reflect the research performance but rather the research outcome of a BuS. Since it is not plausible that the other research performance indicators under consideration must be directly related to citations and, thus, to the research outcome, we assume that correlations between the indicator values are low and, therefore, in the following we will examine whether the citation metrics subsequently considered are valid in terms of discriminant validity. A different picture should emerge if the different citation metrics are compared to each other. All indicators of this type should map the aspect of performance outcome and, therefore, should display high correlations with each other in terms of convergent validity.

2.3 Previous research findings and research deficits

The international research literature contains numerous performance analyses of universities (e.g. Bornmann et al. 2013), departments (e.g. Ketzler and Zimmermann 2013), and individual researchers (e.g. Clarke 2009) based on citation metrics. In these analyses, the performance of the respective organizational unit is either only measured on the basis of citations (e.g. McKercher 2008) or in relation to publication indicators (e.g. Zhu et al. 2014). In contrast, there are only a few studies of a citation analysis of business administration research in Germany. Thus, for example, Sternberg and Litzenberger (2005) analyse the publication- and citation-based research performance of the, at the time, twelve largest university departments for business and social science in Germany. Dyckhoff et al. (2005b) as well as Dyckhoff and Schmitz (2007) also investigate these performance components—but on the level of individual researchers—by focusing on their international visibility. Dilger (2010) as well as Dilger and Müller (2012), on the other hand, generate personal rankings of German business administration researchers on the basis of citations. In contrast to the above-mentioned studies, which emphasize the determination, identification, and analysis of scientists' research performance, Waldkirch et al. (2013) investigate the impact of citation metrics on the performance of German business administration researchers. They focus on individual researchers in the subfields of accounting and marketing.

To the best of our knowledge, only analyses of the relation between citation metrics and publication indicators have been undertaken to date with respect to the relation between research performance indicators.[5] For example, in their empirical analysis of German economics researchers, Schläpfer and Schneider (2010) conclude that there is only a weak correlation between citations and publications so that if citations are not taken into consideration as a performance indicator, important information on the character and quality of research performance of German economists will be lost. Furthermore, Waldkirch et al. (2013) show that there is a difference between those rankings based on numbers of publications and

[5] This must be made distinguished from studies investigating factors influencing certain research performance indicators. Eisend and Schuchert-Güler (2015), for example, analyse the extent to which the composition of a research team with respect to gender and internationality influences the journal publication success. The journal publication success is formalized on the basis of different publication indicators.

those based on citation metrics. With respect to the h-index, on the one hand, and the number of publications, on the other hand, van Raan (2006) notes a rather weak linear relation for Dutch researchers in the field of chemistry, whereas Costas and Bordons (2007) demonstrate a stronger linear relationship with respect to Spanish researchers in the natural sciences. Other empirical studies analyse the relationships between the h-index, its variations and the number of citations and citations per paper. In their metastudy based on 135 correlation coefficients from 32 studies, Bornmann et al. (2011) examine the relationship between the h-index and 37 different h-index variants. They conclude (p. 356): "With an overall mean value between 0.8 and 0.9, there is a high correlation between the h index and the h index variants". These findings are supported by more recent studies (Saxena et al. 2011; Abramo et al. 2013). With respect to German business administration researchers in the subdisciplines of finance and marketing, Breuer (2009) shows high rank correlations between rankings based on citations and on h-indexes. In their study of business researchers in the research fields of accounting and marketing, Waldkirch et al. (2013) also detect high rank correlations between evaluations on the basis of the h-index and on the basis of citations as well as between rankings on the basis of the h-index and the g-index.

It becomes apparent that there is still a need for further research. The presented studies and the resulting empirical findings are only based on correlations between citation metrics and publication indicators. Therefore, it is more or less plausible that closer relationships considering these related indicators are observed. To the best of our knowledge, however, the extent to which there is a relationship between citations and other research performance indicators, for example the level of third-party funding or the number of PhDs, has not yet been analysed.

In selecting the appropriate indicators for measuring research performance, consideration should not be given, or at least not exclusively, to absolute indicators, since this leads to the danger of distortions or undesirable effects (Ursprung and Zimmer 2007). This is because absolute indicators are dependent on both size and discipline (e.g. Zitt et al. 2005). To date, however, analyses of the relations between publication-based and citation-based indicators have primarily been performed on the basis of absolute data. It is, therefore, possible that a size effect is present in the results, since, for example, the more professors perform research at a university, the more they publish, as a rule, and thus at least the probability increases that higher citation rates are achieved. In the present study, we therefore intend to concentrate exclusively on relative indicators.

Investigations of the relations between individual citation metrics have mainly been conducted for other disciplines and/or other countries. However, various authors (e.g. Linton et al. 2011; Albers 2015; Clermont and Dirksen 2016) draw attention to the fact that performance at universities whether between different subject areas or different countries can hardly be compared with each other. For this reason, the results of the above-mentioned studies are not necessarily transferable to German BuSs. Only the analyses by Waldkirch et al. (2013) provide a first starting point. However, Waldkirch et al. focus on the level of individual researchers, particularly on those publishing in the fields of accounting and marketing. The findings are, therefore, not necessarily transferable to BuSs as a whole, because they involve other areas of business administration research.

None of the above-mentioned studies investigated the relationship between research performance indicators in general and citation metrics in particular with journal-based normalizing metrics irrespective of subject or country. Such metrics are intended to cancel out differences in the scholarly communication of different disciplines to achieve comparability across the boundaries of disciplines. In this sense, Colliander and Ahlgren (2011) argue: "By calculating normalized impact (…) and contrasting the result with indicators normalized on a higher aggregated level, a more informative picture can emerge, compared to the case where only one baseline is used. The use of multiple baselines is also valuable for indicating how robust, with respect to the choice of baseline, the rank of a particular unit is." The use of such indicators makes sense if subdisciplines are assumed to cause distortions within one academic discipline (Ball et al. 2009). According to the results of the analysis by Dilger and Müller (2012), such distortions are also found in the individual subdisciplines of business administration.

3 Research design

Table 1 gives an overview of all the research indicators used in this study, their respective definition, the unit of measurement applied, the indicator type, and also details of the indicator source and period. As can be seen from Table 1, in the following we use two publication-based and two supportive indicators taken from the CHE research ranking. Furthermore, five citation-based indicators are generated.

Whether the CHE research ranking can provide a solid basis for a performance analysis of BuSs is confirmed by some authors (Tavenas 2004; Usher and Savino 2006; Marginson and van der Wende 2007; Stolz et al. 2010) but also regarded critically by others (e.g. Ursprung 2003; Ahn et al. 2007; Frey 2007; Clermont and Dirksen 2016). However, in general, CHE provides a comprehensive, regularly surveyed data basis that has been successively further developed since the first survey in the year 2001 and is adequate to serve as a basis for empirical analyses (see, e.g. Ahn et al. 2007; Albers and Bielecki 2012; Bielecki and Albers 2012; Dyckhoff et al. 2013; Clermont et al. 2015). To this end, we used the data collected by CHE in 2010. These data refer to the period from 2007 to 2009 and were made available to us by CHE.[6] A BuS consists of all the departments, institutes and chairs assigned by the university or BuS as business administration units and thus the scientists employed there.[7]

[6] Before utilizing these data, we subjected them to a plausibility check. For example, the data processing performed as part of the ranking was checked on the basis of the raw data provided. In addition to minor corrections, in particular of the number of personnel, above all an error in the internationally visible publications was corrected. 269 international publications were erroneously assigned to LMU Munich. However, only 38 publications could actually be found in the Web of Science database for the period under consideration. We therefore only made use of these corrected data sets in our analyses.

[7] Attention should be drawn to the fact that some German BuSs did not participate in the CHE surveys. The BuS performance of the universities of Cologne and Hamburg, for example, is not evaluated. Since it is not possible to subsequently collect data on the performance of these BuSs in a methodologically stringent manner, in the following we only analyse the performance of those BuSs which participated in the 2010 CHE survey.

Table 1 Set of indicators analysed

Indicator	Definition	Unit of measurement	Indicator type	Database	Data source	Period
Nationally visible publication points per researcher with PhD degree	Number of articles by a researcher in the database WISO, weighted by number of authors, length and type of the article and in the case of journal articles by the quality of the publication medium on the basis of the journal ranking JOURQUAL2 (Schrader and Hennig-Thurau 2009)	Individual researcher	Publication-based	WISO	CHE	2007–2009
Internationally visible publications per researcher	Number of articles of a BuS in the Web of Science databases	BuS	Publication-based	Web of science	FZ Jülich	2007–2009
Amount of third-party funds expended per professor	Amount of third-party funds of a BuS in euros	BuS	Supportive	BuSs	CHE	2007–2009
Number of PhD dissertations per professor	Number of completed PhD dissertations in the semesters of the period under consideration	BuS	Supportive	BuSs	CHE	2007–2009
Averaged citations per paper	Number of citations of all papers of the researchers of a BuS divided by the number of papers	BuS	Citation-based	Web of science	Own computation	2007–2014
h-Index	h of a total of N papers of a BuS have at least h citations, whereas the remaining (N − h) papers have h citations or fewer than h citations	BuS	Citation-based	Web of science	Own computation	2007–2014
e-Index	The citations of a BuS above h correspond to e articles, which would have been cited precisely e times	BuS	Citation-based	Web of science	Own computation	2007–2014
g-Index	With a publication list of a BuS in decreasing order with respect to the number of citations, the g-index is the ranking r in which g publications achieve a citation value of at least r^2	BuS	Citation-based	Web of science	Own computation	2007–2014

Table 1 continued

Indicator	Definition	Unit of measurement	Indicator type	Database	Data source	Period
J-Factor	Relation between the citation frequency of the articles of a BuS in a publication medium or journal and the citation frequency of the articles of a reference group in the same publication medium or journal	BuS	Citation-based	Web of science	Own computation	2007–2014

Indicator type: Publication-based, amounts of the indicators of this type depend primarily on the publications or articles of the unit of measurement; Citation-based, amounts of the indicators of this type depend primarily on the citations of articles of the unit of measurement; Supportive, amounts of the indicators of this type give information about the support of the research or promotion of young researchers of the unit of measurement

BuS Business school, *CHE* Centre for Higher Education, *FZ Jülich* Bibliometrics team of the Central Library of Forschungszentrum Jülich

Since a good (critical) survey of the individual CHE indicators is given in Clermont and Dirksen (2016), we shall dispense with a detailed representation of the indicators here and confine ourselves to the citation metrics. In addition to the number of citations per paper, the h-index (Hirsch 2005) has been extensively discussed and its validity empirically examined (e.g. Hirsch 2007; Jensen et al. 2009; Hönekopp and Khan 2012; Sharma et al. 2013). In a ranking of papers by a BuS D listed in descending order, let r be the rank of a journal in this ranking and $c(r)$ the citations of the paper in the r-th place. Then, the following is valid for calculating the h-index of a BuS:

$$h(D) = \max_{r} \min(r, c(r)). \tag{1}$$

An advantage of this metric is that highly cited papers are not excessively weighted (Rousseau and Leuven 2008). Moreover, the robustness of the h-index with respect to the database and self-citations is highlighted (Vanclay 2007; Hirsch 2005). The basic criticism of the h-index is that the supposed strength of its robustness at the same time constitutes its major weakness, since, for example, individual papers with excellent citation figures in a scientist's publication career receive correspondingly less attention (Harzing and van der Wal 2009). This particularly applies to researchers with only a few papers that are, however, highly cited (Paludkiewicz and Wohlrabe 2010; Bornmann et al. 2011).

Due to this criticism, further developments and variations based on the h-index, which integrate different aspects, such as the number of authors or the age of the article in the calculation of the index (for an overview, see, e.g. Bornmann and Daniel 2009; Harzing 2011), are given far less consideration in the literature.[8] The problem of neglecting papers above the h-index is also taken into account in both the e-index (Zhang 2009) and the g-index (Egghe 2006). The e-index makes use of the results of an h-index calculation. Let c_i be the number of citations of the i-th paper of a BuS D and $h(D)$ the h-index of this BuS from (1), the e-index is then calculated as:

$$e(D) = \sqrt{\sum_{i=1}^{h} c_i - h(D)^2}. \tag{2}$$

The term beneath the square root integrates the additional citations of the h-papers that are not considered in the h-index; the greater its value, the more information that would not be considered was the h-index alone to be applied. The g-index, in contrast, differs from the h-index to the extent that articles with a large number of citations are weighted more heavily. The index value is formed as follows: With a publication list of a university department D in decreasing order with respect to the number of citations c_i, the g-index is the ranking r in which g publications achieve a citation value of at least r^2 in total (Egghe 2006, p. 132):

[8] Furthermore, it should be noted that the original h-index and its variants largely focus on the evaluation of individuals. Many of the variants were developed to compensate for distortions or inaccuracies in the h-index. Since in the present investigations we do not focus on individuals but rather on the level of BuSs, many of the inaccuracies are averaged out due to the enlarged parent population of publications. We have, therefore, not considered supplementary variants such as age or period of scientific activities. Variants that include the age of publications are also irrelevant in our analyses, since the publications by the BuSs investigated are from the same publication years.

$$g(D) = \max\left\{ r | \sum_{i=1}^{r} c_i \geq r^2 \right\}. \quad (3)$$

Harzing and van der Wal (2009) assert that the g-index is a useful complement to the h-index although it has received little attention. However, Zhang (2009) notes that the g-index is not (meaningfully) defined for all possible constellations and in particular not if the sum of the citations of all publications (i.e. of all possible rankings) is greater than the square of the last ranking.

To establish comparability in bibliometric terms between the communication habits of different disciplines or even between subdisciplines within a discipline, we also examine a journal-based normalized citation metric. The J-factor developed by Ball et al. (2009) represents such a metric. It is based on the idea of evaluating the citations c of a unit or group in relation to a predefined, subject-related reference group. That is to say, for each BuS D, the average citations per paper cpp in a certain journal j are related to the average citations per paper in precisely this journal j with the same publication year and the same document type which have been reached by the reference group R (journal-based normalization). This relation is weighted by the proportion of articles p of a BuS in this journal with respect to all the articles n of the BuS during the investigation period. The resulting relations (per BuS and journal) for the respective BuS are then added up over all journals:

$$J(D,R) = \sum_{j=1}^{m} \frac{\mathrm{cpp}_{D,j}}{\mathrm{cpp}_{R,j}} \cdot \frac{p_{D,j}}{n} = \sum_{j=1}^{m} \frac{c_{D,j}/p_{D,j}}{\mathrm{cpp}_{R,j}} \cdot \frac{p_{D,j}}{n} = \frac{1}{n} \sum_{j=1}^{m} \frac{c_{D,j}}{\mathrm{cpp}_{R,j}}. \quad (4)$$

A J-factor of 1 means that the articles of a BuS are cited with exactly the same frequency as those of the reference group. Accordingly, this BuS displays an average citation performance. Correspondingly, J-factors of greater than 1 indicate an above-average citation performance and J-factors of less than 1 a below-average citation performance of the BuS in comparison to the relevant reference group.

To calculate the above-mentioned citation-based indicators, we make use of the internationally visible publications. On the basis of these publication data, we determined the corresponding citations identified in the WoS databases.[9] From a scientometric point of view, however, the CHE survey period of three years is relatively short for a citation analysis. In particular, articles published in 2009 could hardly be perceived, appreciated, assessed, and cited by the scientific community during this time. For this reason, we considerably extend the citation period in comparison to the survey periods of the other CHE indicators. We consider the period from January 2007 to March 2014.[10]

[9] We use the WoS databases, since they are among the most widely utilized interdisciplinary academic literature databases worldwide. In contrast to Scopus, in WoS only the journals with the highest impact in each discipline are indexed. The sources included in WoS primarily focus on the core journals of each discipline. Studies have shown that in the collection of standardized bibliometric metrics there are no significant differences between results based on WoS and on Scopus (e.g. Ball and Tunger 2006b).

[10] The period we selected undoubtedly does not cover all possible citations of a publication. Abramo et al. (2012) comment that: "The accuracy of bibliometric assessment for individual scientists' productivity seems quite acceptable even immediately after a given three-year period and would clearly be even greater for observation periods longer than three years, as typically practiced in national research

In the following, a total of 66 BuSs from Germany will be included in the analysis. Additionally, a reference group has to be defined to calculate the J-factor. For this reference group, we included all those BuSs which were analysed as part of the CHE research performance evaluation published in 2011. CHE's original data basis goes beyond the BuSs mentioned in the CHE publications (see Berghoff et al. 2011) because this data basis includes all BuSs in German-speaking countries, i.e. also those in Austria and Switzerland. To obtain a larger reference group, we also included these BuSs in the reference group.

Correlation analyses have become established as an instrument for investigating relationships between indicators in general and verifying the convergent validity or discriminant validity of indicators in particular. We, therefore, first calculate the two-tailed correlation coefficient of Bravais and Pearson (Pearson 1896), which indicates the approximate level of the linear relationship between two variables. Since our study primarily focuses on an analysis of the influence of using citation metrics and/or an inclusion of citation metrics in an indicator set for research evaluation, we additionally calculate the rank correlation coefficients.[11] These coefficients indicate how the rankings change that are based on the level of attribute variables. In the literature, two methods are primarily discussed in this respect: Spearman's ρ (1904) and Kendall's τ (1938).

The results of a recent study show that Kendall's τ is to be preferred if a small sample is involved and/or if outliers are present in the data set (Xu et al. 2013). In addition, Kendall's τ should be used if the same ranking positions (= bindings) result (Schendera 2004). As shown in the following section, the citation data have a right-skewed distribution with, in places, high outlier values and multiple ranking positions. Furthermore, the volume of data investigated (in spite of an almost complete survey of all existing BuSs in Germany) with 66 BuSs is rather small. Moreover, in contrast to Kendall's τ, Spearman's ρ assumes that differences in the rankings are equivalent. In this sense, equivalence means that, for example, the difference between the best and the second-best BuS is of exactly the same size as the difference between the BuSs in the last and second to last place. However, in the indicators that we consider there is no such equivalence.[12] These are the reasons why we make use of Kendall's τ in the following.

Footnote 10 continued
assessments." In his study of the impact duration of economics publications, Franses (2014) notes that "finally, after 20 years the charts contained 95% new names, and after 10 years it was 85%, suggesting that peak performance rarely lasts for more than 10 years. The most common appearance in the charts is 4 years." Our selected study period is thus of the usual length and should in principle lead to reliable and valid results.

[11] To derive valid findings on the relationship between variables, in the application-oriented literature it is generally recommended that both linear correlations and ranking correlations should be used for an analysis (Hauke and Kossowski 2011).

[12] We would like to thank an anonymous reviewer for this information.

Table 2 Descriptive analysis ($N = 66$ business schools)

	PP-nat per PhD	Pub-int per researcher	Citations per paper	h-Index	e-Index	g-Index	J-Factor	TPF per scientist	PhDs per prof
Arithmetic mean	2.5	0.5	7.3	5.8	7.7	9.6	0.9	24.6	1.1
Median	2.3	0.4	5.9	5.5	6.4	8.0	0.8	18.5	1.0
Minimum	1.2	0.0	0.0	0.0	0.0	0.0	0.0	1.9	0.1
Maximum	4.7	3.2	40.0	18.0	27.8	33.0	1.9	105.5	3.8
Standard deviation	0.8	0.5	7.1	3.9	5.7	7.0	0.4	21.0	0.7
Coefficient of variation	0.3	1.0	1.0	0.7	0.7	0.7	0.4	0.9	0.6

PP-nat per PhD nationally visible publication points per researcher with PhD degree, *Pub-int per researcher* internationally visible publications per researcher, *TPF per researcher* amount of third-party funds expended per researcher, *PhDs per prof* number of PhD dissertations per professor

4 Analysis of the relationships between the different types of research performance indicators

For all research performance indicators taken into consideration, Table 2 displays the arithmetic mean and the median, the minimum and the maximum as well as the standard deviation and the coefficient of variation. With respect to the citations per paper, there is a clear difference between the arithmetic mean and the median, which indicates a positive skew of the citations among the BuSs. This impression is reinforced by looking at the standard deviation and the coefficients of variation, since these are also high for the citations per paper. However, almost all other indicators display a relatively inhomogeneous distribution of the metric values between the BuSs. Lower fluctuations can be observed above all with the nationally visible publication points per researcher with PhD and for the J-factor. Moreover, it is apparent that on average the German BuSs are below the average for the reference group for the J-factor. This observation shows that the German-speaking BuSs which are not located in Germany can be evaluated better with respect to the J-factor than those in Germany.[13]

In Table 3, the correlation coefficients on the relation of the indicators listed in Table 1 are displayed. The first number given in each field of Table 3 corresponds to the correlation coefficient according to Bravais and Pearson, whereas the respective second (lower) number gives the rank correlation coefficient according to Kendall. To identify the level of differences between the significant Bravais/Pearson correlations (see Table 3), the resulting Z values are shown in Table 4. To this end, the correlations were converted in accordance with the Fisher Z-transformation (Dunn and Clark 1969, 1971) and the Z values were calculated (Meng et al. 1992). The significant relationships (above a level of 10%) are shown in boldface in Tables 3 and 4.

[13] This result supports the finding of Storbeck (2012) to the extent that the Handelsblatt's 2012 business administration ranking has Swiss and Austrian BuSs at the top of the ranking list.

Table 3 Correlations and rank correlations between the research performance indicators and the citation metrics

	PP-nat per PhD	Pub-int per researcher	Citations per paper	h-Index	e-Index	g-Index	J-Factor	TPF per researcher	PhDs per prof
PP-nat per PhD	1	0.014	0.077	0.138	0.057	0.094	0.114	0.005	0.175
		0.100	0.034	0.048	0.033	0.030	0.096	−0.036	0.054
Pub-int per researcher		1	0.086	**0.549*****	**0.448*****	**0.566*****	0.070	0.168	0.202
			0.163*	**0.510*****	**0.392*****	**0.482*****	0.082	**0.197****	**0.225****
Citations per paper			1	**0.490*****	**0.833*****	**0.624*****	**0.391*****	0.010	0.038
				0.490***	**0.672*****	**0.584*****	**0.342*****	0.074	0.011
h-Index				1	**0.795*****	**0.940*****	**0.267****	**0.259****	**0.303****
					0.696***	**0.878*****	**0.204****	**0.269*****	**0.248*****
e-Index					1	**0.932*****	**0.348*****	0.096	0.089
						0.836***	**0.269*****	**0.190****	**0.159***
g-Index						1	**0.293****	0.194	0.177
							0.223***	**0.241*****	**0.191****
J-Factor							1	−0.140	−0.048
								−0.107	−0.016
TPF per researcher								1	**0.638*****
									0.358***
PhDs per prof									1

First number: correlation coefficient by Bravais/Pearson. Second number: Rank correlation coefficient by Kendall. $N = 66$ business schools

PP-nat per PhD: nationally visible publication points per researcher with PhD degree, *Pub-int per researcher* internationally visible publications per researcher, *TPF per researcher* amount of third-party funds expended per researcher, *PhDs per prof* number of PhD dissertations per professor

* Significant at 10% level

** Significant at 5% level

*** Significant at 1% level

Table 4 *Z* values to compare the significant correlation coefficient by Bravais/Pearson (Table 3) amongst themselves

Correlation coefficient between		Correlation coefficient between																
		Pub-int &			Citations &				h-index &					e-index &		g-index &	TPF &
		h-index	e-index	g-index	h-index	e-index	g-index	J-factor	e-index	g-index	J-factor	TPF	PhD	g-index	J-factor	J-factor	PhD
Pub-int &	h-index	-	-0.6704	-0.0811	-0.4620	**-3.1977** ***	-0.5984	-1.1594	**-2.6953** ***	**-6.2841** ***	**-1.9168** **	**-1.9771** **	**-1.7335** **	**-5.8370** ***	**-1.4196** *	**-1.7950** **	-0.7846
	e-index		-	-0.9138	-0.2883	**-3.9480** ***	**-1.3487** *	-0.4092	**-3.4456** ***	**-7.0344** ***	-1.1665	-1.2268	-0.9832	**-6.5873** ***	-0.7330	-1.0447	**-1.5349** *
	g-index			-	-0.6256	**-3.0342** ***	-0.4349	**-1.3230** *	**-2.5317** ***	**-6.1206** ***	**-2.0803** **	**-2.1407** **	**-1.8970** **	**-5.6735** ***	**-1.5832** *	**-1.9585** **	-0.6210
Citations &	h-index				-	**-3.6598** ***	-1.0604	-0.6974	**-3.1573** ***	**-6.7461** ***	**-1.4547** *	**-1.5151** *	-1.2715	**-6.2991** ***	-0.9576	**-1.3329** *	-1.2466
	e-index					-	**-2.5993** ***	**-4.3572** ***	-0.5025	**-3.0864** ***	**-5.1145** ***	**-5.1749** ***	**-4.9312** ***	**-2.6393** ***	**-4.6174** ***	**-4.9927** ***	**-2.4132** ***
	g-index						-	**-1.7579** **	**-2.0969** **	**-6.2118** ***	**-2.5152** ***	**-2.5756** ***	**-2.3319** ***	**-5.2386** ***	**-2.0180** **	**-2.3934** ***	-0.1862
	J-factor							-	**-3.8547** ***	**-7.4436** ***	-0.7573	-0.8177	-0.5740	**-6.9965** ***	-0.2602	-0.6355	**-1.9440** **
h-index &	e-index								-	**-3.5888** ***	**-4.6121** ***	**-4.6724** ***	**-4.4288** ***	**-3.1417** ***	**-4.1149** ***	**-4.4902** ***	**-1.9107** **
	g-index									-	**-8.2009** ***	**-8.2612** ***	**-8.0176** ***	-0.4471	**-7.7037** ***	**-8.0791** ***	**-5.4995** ***
	J-factor										-	-0.0604	-0.1833	**-7.7538** ***	-0.4972	-0.1218	**-2.7013** ***
	TPF											-	-0.2436	**-7.8142** ***	-0.5575	-0.1822	**-2.7617** ***
	PhD												-	**-7.5705** ***	-0.3139	-0.0615	**-2.5181** ***
e-index &	g-index													-	**-7.2566** ***	**-7.6320** ***	**-5.0525** ***
	J-factor														-	-0.3753	**-2.2042** **
g-index &	J-factor															-	**-2.5795** ***
TPF &	PhD																-

Pub-int Internationally visible publications per researcher
Citations Citations per paper
TPF Amount of third-party funds expended per researcher
PhD Number of PhD dissertations per professor

N = 66 business schools
*** Significant at 1% level
** Significant at 5% level
* Significant at 10% level

If the relationship between the publication-based and the citation-based indicators is first considered, then it can be seen that there are no significant relations with respect to the national publication points per researcher with PhD. Nor can any significant relationship be established between the international publications per researcher and the J-factor, whereas there is a low rank correlation with the citations per paper. With respect to the three citation index indicators, the international publications per researcher display positive relationships at a medium level, which does not, however, differ with respect to the Bravais/Pearson correlation (see Table 4). These results thus vary from those in Sect. 2.3. The majority of studies mentioned there show a strong positive relation between the publication-based and the citation-based indicators. In these studies, however, mainly absolute values, such as total number of publications and total number of citations, were used as reference units whereas in our investigation we used relative measures. If we consider absolute measures, then our data also reveal considerably stronger relations. For example, there is a linear relationship of 0.449 or 0.745 between the absolute number of citations and the national publication points or international publications. This suggests that a size effect occurs here in the form of the Matthew effect (Merton 1968).

Only slight relationships can be established between the two supportive and citation-based indicators, and are generally restricted to weak rank correlations. A weak linear relationship is only found between third-party funds or PhDs and the h-index. It is striking that if the two supportive indicators are considered, the respective relationships are at a similar (low) level. This may be due to the fact that both indicators display an above-average significant linear relationship of 0.638 among each other, which is not surprising with respect to their topic. In business administration, it is largely staff positions that are funded by third-party funds, supplementing those positions financed by basic government funding. The primary task of these externally funded staff is research. At the same time, these positions are usually linked to work on a PhD.

It remains to take a look at the relations of the citation-based indicators to each other. In contrast to the above-mentioned results, entirely positive significant relationships between the individual indicators are found here. In agreement with previous empirical results in the literature for other disciplines and/or countries, it can be determined that the h-index and its variants do not lead to any great differences in the respective rankings. Highly significant correlations between the h-index and the g-index become obvious, and to a lesser extent between the h- and the e-index. There are thus indications that with rising h-index the additional citations above the g-index also increase in a linear relationship. The citations per paper display low to medium correlations with all other citation metrics. Interestingly, in the case of the e-index a relatively high correlation can be identified as 0.833 or 0.672. With respect to all other citation metrics, the J-factor displays rather low correlations, whose differences are, moreover, not significant.

5 Analysis of top groups for indicator values

In this section, we first analyse the difference in the composition of the so-called top BuSs with respect to the research performance indicators and the citation metrics. We applied the CHE procedure for relative indicators to create the top groups, i.e. the best 25% of the BuSs form the top group. If two or more BuSs with identical indicator values are found at the transition of the top group, then all these BuSs are allocated to the top group, which, as in the present case, leads to slightly different group sizes among the partial ratings. The resulting top groups are shown in alphabetical order in Table 5.

We conducted Chi square tests to investigate the extent to which the individual partial ratings differ. Since, on the one hand, we only considered ratings and not rankings, and, on the other hand, different BuSs were present in the top groups, we undertook a binary consideration by assigning each BuS in a top group the value "1" and, correspondingly, all other BuSs were assigned the value "0". The Chi square values generated on this basis are given in Table 6 together with the related significance level.

A consideration of the top groups reveals close correlations between the ratings of the top groups formed on the basis of citation metrics, especially for the citation index indicators. However, the ratings drawn up on the basis of the citations per paper also display considerable parallels with all other citation ratings. With the exception of the internationally visible publications per researcher and the TPF per researcher, hardly any correlations can be established between the top groups of the other indicator types. The results obtained in Sect. 4 are thus also corroborated with respect to the top groups.

We finally consider whether BuSs classified as strong in research or reputable can be identified in the partial ratings of the individual indicators. Therefore, we make use of three studies: the investigation of BuSs strong in research by CHE and Handelsblatt as well as the survey of reputable BuSs by CHE. According to CHE, a BuSs is designated strong in research if it belongs to the top group in four out of eight partial ratings of different research performance indicators. For partial ratings on the basis of absolute figures, the top group is determined by those BuSs with the highest values which generate a cumulative 50% of the total sum of the respective indicator values. In the four relative performance ratings, the 25% best BuSs are included in the top group. To counteract any distortions due to the data collection methods, we considered all BuSs as being strong in research which had been among the top groups at least twice in the last three surveys.

The classification strong in research by Handelsblatt resulted on the basis of the number of journal papers published by the researchers. Various journal ratings and rankings are used for the quality weighting of the journals, whereby the individual papers are assigned between zero and one point. In addition, the number of co-authors of a paper is considered by the Handelsblatt and the number of points awarded is divided by this number. Since Handelsblatt also included Austrian and Swiss BuSs, we will only consider the 18 BuSs in Germany from the ranking of the TOP 25 BuSs.

Table 5 Top groups in partial ratings of the indicators under consideration

PP-nat per PhD	Pub-int per researcher	Citations per paper	h-Index	e-Index	g-Index	J-Factor	TPF per researcher	PhDs per prof
Bamberg	Berlin FU	Bamberg	Aachen	Aachen	Aachen	Augsburg	Aachen	Aachen
Berlin ESCP-EAP	Berlin HU	Berlin ESCP-EAP	Bamberg	Berlin FU	Berlin FU	Berlin ESCP-EAP	Berlin HU	Augsburg
Berlin TU	Berlin TU	Berlin FU	Berlin FU	Berlin HU	Berlin HU	Berlin FU	Berlin TU	Berlin TU
Bochum	Bielefeld	Berlin HU	Berlin TU	Bielefeld	Berlin TU	Berlin HU	Bremen	Bremen
Bremen	Bremen	Bielefeld	Bielefeld	Dresden	Bielefeld	Duisburg	Chemnitz	Chemnitz
Düsseldorf	Duisburg	Duisburg	Bremen	Duisburg	Bremen	Göttingen	Dresden	Erlangen-Nürnberg
Frankfurt Oder	Flensburg	Hamburg Uni	Duisburg	Frankfurt Main Uni	Dresden	Hagen	Erlangen-Nürnberg	Flensburg
Hohenheim	Frankfurt Main Uni	Hamburg UBW	Frankfurt Main Uni	Hamburg Uni	Duisburg	Halle-Wittenberg	Flensburg	Hannover
Jena	Frankfurt Oder	Jena	Hamburg Uni	Hannover	Frankfurt Main Uni	Hamburg UBW	Frankfurt Main SFM	Hohenheim
Kiel	Hamburg Uni	Kassel	Hannover	Jena	Hamburg Uni	Ilmenau	Frankfurt Main Uni	Kassel
Leipzig HH	Hannover	Konstanz	Jena	Kassel	Hannover	Jena	Greifswald	Mannheim
München LMU	München LMU	Magdeburg	Mannheim	Magdeburg	Jena	Kassel	Kaiserslautern	München LMU
München TU	MünchenTU	Mannheim	München LMU	Mannheim	Magdeburg	Konstanz	München TU	München TU
München UBW	Oestrich-Winkel	Marburg	München TU	München LMU	Mannheim	Marburg	Oestrich-Winkel	Münster
Saarbrücken	Potsdam	München LMU	Oestrich-Winkel	München TU	München LMU	München LMU	Oldenburg	Oestrich-Winkel
Stuttgart	Siegen	München UBW	Oldenburg	Oldenburg	München TU	München UBW	Potsdam	Regensburg

Table 5 continued

PP-nat per PhD	Pub-int per researcher	Citations per paper	h-Index	e-Index	g-Index	J-Factor	TPF per researcher	PhDs per prof
Vallendar	Tübingen	Oldenburg	Vallendar	Vallendar	Oestrich-Winkel	Rostock	Vallendar	Stuttgart
	Ulm	Vallendar			Oldenburg	Stuttgart		Trier
					Vallendar	Vallendar		Witten-Herdecke
						Würzburg		

PP-nat per PhD nationally visible publication points per researcher with PhD degree, *Pub-int per researcher* internationally visible publications per researcher, *TPF per researcher* third-party funds expended per researcher, *PhDs per prof* PhD dissertations per professor, *ESCP-EAP* Ecole Supérieure de Commerce de Paris-Ecole des Affaires de Paris, *FU* Free University, *HH* Commercial College, *HU* Humboldt University, *LMU* Ludwig Maximilian University, *SFM* School of Finance and Management, *TU* University of Technology, *UBW* University of the German Armed Forces, *Uni* University

Table 6 Chi square tests to compare the top groups

	PP-nat per PhD	Pub-int per researcher	Citations per paper	h-Index	e-Index	g-Index	J-Factor	TPF per researcher	PhDs per prof
PP-nat per PhD	–	0.530	1.089	**2.847***	**0.160**	0.473	0.270	0.059	0.473
Pub-int per researcher		–	0.053	**16.176*****	**7.606*****	**17.322*****	0.765	**4.520****	1.232
Citations per paper			–	**13.092*****	**13.092*****	**10.076*****	**12.832*****	2.345	1.386
h-Index				–	**30.795*****	**47.671*****	0.009	**5.433****	**3.729***
e-Index					–	**39.473*****	1.282	**2.847***	0.473
g-Index						–	0.021	**10.076*****	2.308
J-Factor							–	**3.726***	1.081
TPF per researcher								–	**3.729***
PhDs per prof									–

Bold value indicates significant correlations

N = 66 business schools

PP-nat per PhD nationally visible publication points per researcher with PhD degree, *Pub-int per researcher* internationally visible publications per researcher, *TPF per researcher* amount of third-party funds expended per researcher, *PhDs per prof* number of PhD dissertations

* Significant at 10% level

** Significant at 5% level

*** Significant at 1% level

Table 7 Percentage of strong in research and reputable BuSs in the respective partial ratings

	Strong in research (CHE) $N = 16$ (%)	Strong in research (Handelsblatt) $N = 15$ (%)	Reputable $N = 9$ (%)
PP-nat per PhD	31	27	44
Pub-int per researcher	44	40	56
Citations per paper	31	47	56
h-Index	63	67	67
e-Index	56	67	78
g-Index	63	67	78
J-Factor	25	33	44
TPF per researcher	50	40	44
PhDs per prof	69	47	44

The classification of the research reputation of a BuS was determined on the basis of a supplemental CHE survey. CHE requested all professors of business administration in Germany in each data acquisition year to state up to 5 BuSs that they perceived as being notably reputable. Professors were not permitted to nominate their own BuS. Again, in order to counteract any data distortions, we considered all BuSs as being reputable which had been among the top 10 BuS at least twice in the last three reputation surveys.

Finally, we filtered out all those BuSs from the generated lists of BuSs strong in research or reputable which were not in our data set of 66 BuSs. We then ultimately obtained 16 BuSs classified as "strong in research" (according to CHE), 15 BuSs as "strong in research" (according to Handelsblatt), and 9 BuSs as "reputable". In the following Table 7, we show the proportion of those BuSs classified as strong in research or reputable which can be identified in the respective top groups.

It becomes apparent that more than 50% of the CHE strong in research BuSs are represented in the partial ratings of the h-, g-, and e-indexes as well as in the partial ratings of the two supportive indicators. With respect to the Handelsblatt classification, more than 50% of these BuSs can be found in the three above-mentioned citation index indicators. The reputable BuSs are also represented in all three citation index indicators with over 67%. In addition, slightly more than 50% of the reputable BuSs are listed in the partial ratings of the international research publications per researcher and the citations per paper. With respect to the citation metrics, strong relations emerge between the top 25% rankings for the h-, g-, and e-index and all three BuSs classifications taken into consideration. It is not particularly surprising to discover that this is the case for all three partial ratings of these citation index indicators. Nevertheless, the large number of BuSs either strong in research or reputable is astonishing. For the citations per paper, and especially for the J-factor, however, a different picture emerges, as there are only a few of the classified BuSs in these two partial ratings.

6 Implications, limitations and outlook

6.1 Implications of the study

Within the framework of this study, we posed the question of what relationships exist between citation-based, publication-based, and supportive indicators for representing research performance. To generate corresponding results, we use the data from the CHE, which provide information on the research performance of BuSs in Germany. The results are used to analyse the validity of citation metrics. As already shown, the question of high or low correlations combined with the validity criteria decisively depends on whether the two indicators considered are measures for the same construct or different constructs. In the latter case, for valid indicators the correlations should remain at a low level (divergent validity), whereas in the former case they should be at a high level (convergent validity).

We assume that publication-based and supportive indicators describe the activity level of a BuS and are classically regarded as indicators for mapping research performance. In contrast, citation metrics describe the perception or the impact of publication activities. For this reason, citation metrics are not indicators mapping research performance but rather research outcome. Citation metrics should, therefore, be divergently valid and there should not be any or only low correlations. Our analyses reveal these expectations, because there are either no relationships between the relatively defined citation-based, publication-based, and supportive indicators, or these relations are of low significance.

Interestingly, we were also unable to establish any or only low correlations between the publication-based and the supportive indicators although they are all intended to map the construct of research performance. However, especially in the case of the supportive indicators, it is a matter of debate whether they actually map research performance. We use the example of third-party funding to illustrate that its classification as a research performance indicator is a matter of controversy in the academic literature.[14] This heterogeneous perspective is justified by the fact that, on the one hand, third-party funding is a resource (that should be minimized) and accordingly represents input (see, e.g. Albers 2015), whereas, on the other hand, as a proxy indicator this funding may be associated with the acquisition of money or, however, its use may be anticipated as the predicted desirable research performance (see, e.g. Clermont et al. 2015).

If the citation metrics selected by us map the construct of research outcome, there must be a high correlation between the citation metrics in terms of convergent validity. Indeed high linear correlations and rank correlations can be recognized here, above all between the citation metrics of the h-, g-, and e-indexes. Since journal-based normalized citation metrics are increasingly being discussed in the scientometric literature, we have integrated such an indicator in the form of the J-factor. However, we were only able to identify low significant relationships with the other citation-based indicators.

[14] A detailed discussion of this issue can be found in Hornbostel (2001) or Rassenhövel and Dyckhoff (2006).

It seems possible to explain this in principle with respect to the J-factor's content. The J-factor is indeed primarily a citation-based indicator. However, due to the normalization of the citations on the basis of an expected value in the individual journals in which the articles were published, not only do the citations received influence the indicator value but also the relation of the citation rate of the published papers to the citation rate of all papers in the same journal in the same year of publication. In the case of the J-factor, the ratio of citations per paper to the expected value is decisive. The opportunity for manipulation is thus considerably reduced since in this consideration it is not only the perception of a paper that plays an important role but also the performance of the community (Ball et al. 2009).

An investigation of the top groups of the partial rankings created on the basis of the individual indicators confirms the above-mentioned study results. As an additional criterion, we compared these partial ratings with the classification of BuSs as strong in research or reputable by CHE and Handelsblatt. It was found that a high proportion of BuSs classified in this way were included in the top 25% of the three citation index indicators. In particular, a large proportion of the BuSs classified as reputable are included in the partial ratings of the citation metrics. With respect to universities, the concept of reputation subsumes the repute, image, prestige, standing, rank, and position of scientists, faculties, or superordinate organizational units (Brenzikofer and Staffelbach 2003). Generally, reputation arises from previous (research) performance (e.g. Cyrenne and Grant 2009; Linton et al. 2011). Albers and Bielecki (2012), for example, showed that the reputation of German BuSs depends on their publication performance. Reputation is, therefore, established as an impact of research performance, which could explain the greater correlation with citation metrics.

6.2 Limitations and outlook for further research

With respect to the survey of citations with literature databases, it should be noted that the resulting citation list is always a more or less partial consideration of the citations of a BuS. Only a certain proportion of the numerous journals relevant for the business administration research community in Germany are covered in the databases of the WoS and the corresponding citations included in them (Clermont and Schmitz 2008; Clermont and Dyckhoff 2012). However, with Scopus and Google Scholar there are not so many alternatives. WoS is regarded in the scientometric community as a particularly popular and reliable data source due to its long-standing presence on the market (Ball and Tunger 2006a). Additionally, Waldkirch et al. (2013) establish a fairly robust correlation between citation metrics for researchers from the fields of accounting and marketing with respect to the specific choice of a literature database.

The J-factor, on the one hand, reacts sensitively to uncited articles and, on the other hand, to cited papers in a journal in which only one paper is included in the entire sample. Given these constellations, the BuS naturally displays an average citation performance for this journal and is assigned a partial J-factor of 100%. Due to the equal weighting of all the partial J-factors, the overall J-factor for a BuS ceteris paribus is closer to 100% the greater the number of such constellations is

present. With respect to uncited papers, this aspect can be moderated by extending the period during which the citations are evaluated. Since, if the citation survey period is extended, then the probability of uncited articles from the same publication period is significantly lower. However, there are a large number of different journals in the field of business administration. JOURQUAL 2.1 of the German Academic Association for Business Research (Schrader and Hennig-Thurau 2009) lists 838 such journals and the 2014 Handelsblatt Journal Rating for business administration lists 761 relevant journals. This increases the probability of a journal containing exclusively articles from the reference group, especially in the case of specialist journals, in which only a few researchers publish their findings. Therefore, it is recommended that a longer survey period should be selected to generate valid and meaningful findings.

It should be noted that there is in principle no recognized ranking of German BuSs. The approach adopted by CHE and Handelsblatt is not universally accepted and has already been subjected to critical scrutiny (cf. e.g. Clermont and Dirksen 2016; Lorenz and Löffler 2015; Müller 2010). Due to their partial (CHE) or exclusive (Handelsblatt) use of absolute indicators, which do not consider the number of staff at the respective BuSs, the larger BuSs have a certain advantage. The size effect could also play a part in the reputation survey because the professors interviewed have a stronger perception of larger BuSs in comparison to smaller ones (see, e.g. De Filippo et al. 2012; Porter and Toutkoushian 2006). The low agreement between these classifications and the relative indicators on which this study focuses could, therefore, also be due to the change of perspective from absolute to relative indicators.

The above-mentioned limitations yield starting points and opportunities for further research. To generate a broader data basis and to give due consideration to the prevailing trend towards internationalization in science, one suggestion would be to enlarge the reference group when calculating the J-factor. In this context, the impact of business administration research in Germany could, for example, be compared to that in Europe, the USA or worldwide. In particular, it may be appropriate to study the impact on the J-factor of an expansion of the reference group on the types of validity. Furthermore, our cross-sectional analyses could be reproduced and expanded as part of longitudinal analyses for future CHE research data acquisition to validate the findings generated here and to identify potential development trends.

Moreover, the question arises of whether the relationships established between different types of indicators for mapping research performance are a special feature of BuSs in German-speaking countries or whether this can also be transferred to other academic disciplines in Germany and/or to other countries. At least for Germany, such studies could also be performed on the basis of the CHE data, since CHE has undertaken research rankings for most academic disciplines, and these rankings represent a solid data basis in principle. In addition, it is also possible to extend the types of performance under consideration since within the framework of this paper we only investigated the relationships between certain aspects of research performance, i.e. aspects of the teaching performance of BuSs were completely neglected.

Acknowledgements The project underlying this article was funded by the Federal Ministry for Education and Research (Germany) under Grant Number 01PW11014. The responsibility for the contents of the article lies with the authors. The authors would like to express their gratitude to the editor and two anonymous reviewers for their constructive comments which contributed to the improvement of the paper. The authors would also thank the Sprachendienst of Forschungszentrum Jülich as well as Ms. Christine Stibbe (RWTH Aachen University) for language editing.

References

Abramo, Giovanni, Tindaro Cicero, and Ciriaco Andrea D'Angel. 2012. A sensitivity analysis of researchers' productivity rankings to the time of citation observation. *Journal of Informetrics* 6: 192–201.

Abramo, Giovanni, Ciriaco Andrea D'Angelo, and Fulvio Viel. 2013. The suitability of h and g indexes for measuring the research performance of institutions. *Scientometrics* 97: 555–570.

Ahn, Heinz, and Harald Dyckhoff. 2004. Zum Kern des Controllings: Von der Rationalitätssicherung zur Effektivitäts- und Effizienzsicherung. In *Controlling: Theorien und Konzeptionen*, ed. Ewald Scherm, and Gotthard Pietsch, 501–528. München: Vahlen.

Ahn, Heinz, Harald Dyckhoff, and Roland Gilles. 2007. Datenaggregation zur Leistungsbeurteilung durch Ranking: Vergleich der CHE- und DEA-Methodik sowie Ableitung eines Kompromissansatzes. *Zeitschrift für Betriebswirtschaft* 77: 615–643.

Ahn, Heinz, Marcel Clermont, Harald Dyckhoff, and Yvonne Höfer-Diehl. 2012. Entscheidungsanalytische Strukturierung fundamentaler Studienziele: Generische Zielhierarchie und Fallstudie. *Zeitschrift für Betriebswirtschaft* 82: 1229–1257.

Albers, Sönke. 2015. What Drives Publication Productivity in German Business Faculties? *Schmalenbach Business Review* 67: 6–33.

Albers, Sönke, and Andre Bielecki. 2012. *Wovon hängt die Leistung in Forschung und Lehre ab? Eine Analyse deutscher betriebswirtschaftlicher Fachbereiche basierend auf den Daten des Centrums für Hochschulentwicklung*. Kiel: Universität Kiel.

Ball, Rafael, Bernhard Mittermaier, and Dirk Tunger. 2009. Creation of journal-based publication profiles of scientific institutions: A methodology for the interdisciplinary comparison of scientific researcher based on the J-factor. *Scientometrics* 81: 381–392.

Ball, Rafael, and Dirk Tunger. 2006a. Bibliometric analysis: A new business area for information professionals in libraries? *Scientometrics* 66: 561–577.

Ball, Rafael, and Dirk Tunger. 2006b. Science indicators revisited: Science citation index versus SCOPUS: A bibliometric comparison of both citation databases. *Information Services and Use* 26: 293–301.

Berghoff, Sonja, Petra Giebisch, Cort-Denis Hachmeister, Britta Hoffmann-Kobert, Mareike Hennings, and Frank Ziegele. 2011. *Vielfältige Exzellenz 2011: Forschung, Anwendungsbezug, Internationalität, Studierendenorientierung im CHE Ranking*. Gütersloh: Centrum für Hochschulentwicklung.

Bielecki, Andre, and Sönke Albers. 2012. *Eine Analyse der Forschungseffizienz deutscher betriebswirtschaftlicher Fachbereiche basierend auf den Daten des Centrums für Hochschulentwicklung (CHE)*. Kiel: Universität Kiel.

Brenzikofer, Barbara, and Bruno Staffelbach. 2003. Reputation von Professoren als Führungsmittel in Universitäten. In *Hochschulreformen in Europa—konkret: Österreichs Universitäten auf dem Weg vom Gesetz zur Realität*, ed. Stefan Titscher, and Sigurd Höllinger, 183–208. Wiesbaden: Springer.

Bornmann, Lutz, and Hans-Dieter Daniel. 2009. The state of h index research. *EMBO Reports* 10: 2–6.

Bornmann, Lutz, Rüdiger Mutz, Sven E. Hug, and Hans-Dieter Daniel. 2011. A meta-analysis of studies reporting correlations between the h index and 37 different h index variants. *Journal of Informetrics* 5: 346–359.

Bornmann, Lutz, Rüdiger Mutz, and Hans-Dieter Daniel. 2013. Multilevel-statistical reformulation of citation-based university rankings: The Leiden ranking 2011/12. *Journal of the American Society for Information Science and Technology* 64: 1649–1658.

Bort, Suleika, and Simone Schiller-Merkens. 2010. Publish or perish. *Zeitschrift Führung und Organisation* 79: 340–346.

Breuer, Wolfgang. 2009. Google Scholar as a means for quantitative evaluation of German research output in Business Administration: Some preliminary results. SSRN. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1280033. Accessed 10 Aug 2015.

Campbell, Donald T., and Donald W. Fiske. 1959. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin* 56: 81–105.

Chalmers, Alan Francis. 1990. *Science and its fabrication*. Buckingham: Open University Press.

Clarke, Roger. 2009. A citation analysis of Australian information systems researchers: Towards a new era? *Australian Journal of Information Systems* 15 (2): 23–44.

Clermont, Marcel, and Alexander Dirksen. 2016. The measurement, evaluation, and publication of performance in higher education: An analysis of the CHE research ranking of Business Schools in Germany from an accounting perspective. *Public Administration Quarterly* 40 (2): 133–178.

Clermont, Marcel, Alexander Dirksen, and Harald Dyckhoff. 2015. Returns to scale of Business Administration research in Germany. *Scientometrics* 103: 583–614.

Clermont, Marcel, and Harald Dyckhoff. 2012. Erfassung betriebswirtschaftlich relevanter Zeitschriften in Literaturdatenbanken. *Betriebswirtschaftliche Forschung und Praxis* 64: 325–347.

Clermont, Marcel, and Christian Schmitz. 2008. Erfassung betriebswirtschaftlich relevanter Zeitschriften in den ISI-Datenbanken sowie der Scopus-Datenbank. *Zeitschrift für Betriebswirtschaft* 78: 987–1010.

Colliander, Cristian, and Per Ahlgren. 2011. The effects and their stability of field normalization baseline on relative performance with respect to citation impact: A case study of 20 natural science departments. *Journal of Informetrics* 5: 101–113.

Costas, Rodrigo, and María Bordons. 2007. The h-index: Advantages, limitations and its relation with older bibliometric indicators at the micro level. *Journal of Informetrics* 1: 193–203.

Cronbach, Lee J., and Paul E. Meehl. 1955. Construct validity in psychological tests. *Psychological Bulletin* 52: 281–302.

Cyrenne, Philippe, and Hugh Grant. 2009. University decision making and prestige: An empirical study. *Economics of Education Review* 28: 237–248.

De Filippo, Daniela, Fernando Casani, Carlos García-Zorita, Preiddy Efraín-García, and Elías Sanz-Casado. 2012. Visibility in international rankings: Strategies for enhancing the competitiveness of Spanish universities. *Scientometrics* 93: 949–966.

Delgado-Márquez, Blanca L., M. Ángeles Escudero-Torres, and Nuria E. Hurtado-Torres. 2013. Being highly internationalised strengthens your reputation: An empirical investigation of top higher education institutions. *Higher Education* 66: 619–633.

Dilger, Alexander. 2000. Plädoyer für einen Sozialwissenschaftlichen Zitationsindex. *Die Betriebswirtschaft* 60: 473–485.

Dilger, Alexander. 2010. Rankings von Zeitschriften und Personen in der BWL. *Zeitschrift für Management* 5: 91–102.

Dilger, Alexander, and Harry Müller. 2012. Ein Forschungsleistungsranking auf der Grundlage von Google Scholar. *Zeitschrift für Betriebswirtschaft* 82: 1089–1105.

Dunn, Olive Jean, and Virginia Clark. 1969. Correlation coefficients measured on the same individuals. *Journal of the American Statistical Association* 64: 366–377.

Dunn, Olive Jean, and Virginia Clark. 1971. Comparison of tests of the equality of dependent correlation coefficients. *Journal of the American Statistical Association* 66: 904–908.

Dyckhoff, Harald, Marcel Clermont, Alexander Dirksen, and Eleazar Mbock. 2013. Measuring balanced effectiveness and efficiency of German business schools' research performance. *Zeitschrift für Betriebswirtschaft Special Issue* 3 (2013): 39–60.

Dyckhoff, Harald, Sylvia Rassenhövel, Roland Gilles, and Christian Schmitz. 2005a. Beurteilung der Forschungsleistung und des CHE-Forschungsrankings betriebswirtschaftlicher Fachbereiche. *Wirtschaftswissenschaftliches Studium* 34: 62–69.

Dyckhoff, Harald, and Christian Schmitz. 2007. Forschungsleistungsmessung mittels SSCI oder SCI-X? Internationale Sichtbarkeit und Wahrnehmung der Betriebswirtschaftslehre von 1990 bis 2004. *Die Betriebswirtschaft* 67: 640–664.

Dyckhoff, Harald, Annegret Thieme, and Christian Schmitz. 2005b. Die Wahrnehmung deutschsprachiger Hochschullehrer für Betriebswirtschaft in der internationalen Forschung: Eine Pilotstudie zu Zitationsverhalten und möglichen Einflussfaktoren. *Die Betriebswirtschaft* 65: 350–372.

Egghe, Leo. 2006. Theory and practice of the g-Index. *Scientometrics* 69: 131–152.

Eisend, Martin. 2011. Is VHB-JOURQUAL2 a good measure of scientific quality? Assessing the validity of the major business journal ranking in German-speaking countries. *Business Research* 4: 241–274.

Eisend, Martin, and Pakize Schuchert-Güler. 2015. Journal publication success of German business researchers: Does gender composition and internationality of the author team matter? *Business Research* 8: 171–188.

Federal Ministry of Education and Research. 2010. Bekanntmachung des Bundesministeriums für Bildung und Forschung von Richtlinien zur Förderung von Forschungsvorhaben zum Themenfeld „Wissenschaftsökonomie“vom 13. April 2010. http://www.bmbf.de/foerderungen/14677.php. Accessed 29 Dec 2015.

Franses, Philip Hans. 2014. Trends in three decades of rankings of Dutch economists. *Scientometrics* 98: 1257–1268.

Frey, Bruno S. 2007. Evaluierungen, Evaluierungen … Evaluitis. *Perspektiven der Wirtschaftspolitik* 8: 207–220.

Harzing, Anne-Wil, and Ron van der Wal. 2009. A Google Scholar h-index for journals: An alternative metric to measure journal impact in economics and business. *Journal of the American Society for Information Science and Technology* 60: 41–46.

Harzing, Anne-Wil. 2011. *The publish or perish book: Your guide to effective and responsible citation analysis*. Melbourne: Tarma Software Research Pty Ltd.

Hauke, Jan, and Tomasz Kossowski. 2011. Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data. *Quaestiones Geographicae* 30: 87–93.

Hirsch, Jorge E. 2005. An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America* 102: 16569–16572.

Hirsch, Jorge E. 2007. Does the h index have predictive power? *Proceedings of the National Academy of Sciences* 104: 19193–19198.

Hönekopp, Johannes, and Julie Khan. 2012. Future publication success in science is better predicted by traditional measures than by the h index. *Scientometrics* 90: 843–853.

Hornbostel, Stefan. 2001. Third party funding of German universities: An indicator of research activity? *Scientometrics* 50: 523–537.

Iñiguez De Onzoño, Santiago, and Salvador Carmona. 2007. The changing business model of B-schools. *Journal of Management Development* 26: 22–32.

Jensen, Pablo, Jean-Baptiste Rouquier, and Yves Croissant. 2009. Testing bibliometric indicators by their prediction of scientists promotions. *Scientometrics* 78: 467–479.

Keeney, Ralph L., and Howard Raiffa. 1993. *Decisions with multiple objectives: Preferences and value tradeoffs*, 2nd ed. Cambridge: Cambridge University Press.

Kendall, Maurice G. 1938. A new measure of rank correlation. *Biometrika* 30: 81–93.

Ketzler, Rolf, and Klaus F. Zimmermann. 2013. A citation-analysis of economic research institutes. *Scientometrics* 95: 1095–1112.

Kleiner, Matthias. 2010. *„Qualität statt Quantität“: Neue Regeln für Publikationsangaben in Förderanträgen und Abschlussberichten*. Statement by the president of the German Research Foundation (DFG) at a press conference on 23 October 2010.

Kromrey, Helmut, and Jochen Roose. 2016. *Empirische Sozialforschung: Modelle und Methoden der standardisierten Datenauswertung*. 13th edition. Stuttgart: utb.

Linton, Jonathan D., Robert Tierney, and Steven T. Walsh. 2011. Publish or perish: How are research and reputation related? *Serials Review* 37: 244–257.

Lorenz, Daniela, and Andreas Löffler. 2015. Robustness of personal rankings: The Handelsblatt example. *Business Research* 8: 189–212.

McKercher, Bob. 2008. A citation analysis of tourism scholars. *Tourism Management* 29: 1226–1232.

MacRoberts, Michael H., and Barbara R. MacRoberts. 1996. Problems in citation analysis. *Scientometrics* 36: 435–444.

Marginson, Simon, and Marijk van der Wende. 2007. To rank or to be ranked: The impact of global rankings in higher education. *Journal of Studies in International Education* 11: 206–329.

Meng, Xiao-Li, Robert Rosenthal, and Donald B. Rubin. 1992. Comparing correlated correlation coefficient. *Psychological Bulletin* 1: 172–175.

Merton, Robert K. 1968. The Matthew effect in science. *Science* 159: 56–63.

Messick, Samuel. 1995. Standards of validity and the validity of standards in performance assessment. *Educational Measurement—Issues and Practice* 14(4): 5–8.

Müller, Harry. 2010. Wie valide ist das Handelsblatt-BWL-Ranking? Zeitschriften- und zitationsbasierte Personenrankings im Vergleich. *Betriebswirtschaftliche Forschung und Praxis* 62: 150–164.

Müller, Harry. 2012. Zitationen als Grundlage von Forschungsleistungsrankings: Konzeptionelle Überlegungen am Beispiel der Betriebswirtschaftslehre. *Beiträge zur Hochschulforschung* 34: 68–92.

Neely, Andy, Mike Gregory, and Ken Platts. 1995. Performance measurement system design: A literature review and research agenda. *International Journal of Operations & Production Management* 15: 80–116.

Nosek, Brian A., Jesse Graham, Nicole M. Lindner, Selin Kesebir, Carlee Beth Hawkins, Cheryl Hahn, Kathleen Schmidt, Matt Motyl, Jennifer Joy-Gaba, Rebecca Frazier, and Elizabeth R. Tenney. 2010. Cumulative and career-stage citation impact of social-personality psychology programs and their members. *Personality and Social Psychology Bulletin* 36: 1283–1300.

Palomares-Montero, Davinia, and Adela García-Aracil. 2011. What are the key indicators for evaluating the activities of universities? *Research Evaluation* 20: 353–363.

Paludkiewicz, Karol, and Klaus Wohlrabe. 2010. Qualitätsanalyse von Zeitschriften in den Wirtschaftswissenschaften: Über Zitationsdatenbanken und Impaktfaktoren im Online-Zeitalter. *Ifo Schnelldienst* 63 (21): 18–28.

Pearson, Karl. 1896. Mathematical contributions to the theory of evolution III: Regression, heredity, and panmixia. *Philosophical Transactions of the Royal Society Series A* 187: 253–318.

Porter, Stephen R., and Robert K. Toutkoushian. 2006. Institutional research productivity and the connection to average student quality and overall reputation. *Economics of Education Review* 25: 605–617.

Rassenhövel, Sylvia, and Harald Dyckhoff. 2006. Die Relevanz von Drittmittelindikatoren bei der Beurteilung der Forschungsleistung im Hochschulbereich. In *Fortschritt in den Wirtschaftswissenschaften: Wissenschaftstheoretische Grundlagen und exemplarische Anwendungen*, ed. Stephan Zelewski, and Naciye Akca, 85–112. Wiesbaden: Gabler.

Rassenhövel, Sylvia. 2010. *Performancemessung im Hochschulbereich: Theoretische Grundlagen und empirische Befunde*. Wiesbaden: Gabler.

Rousseau, Ronald, and K.U. Leuven. 2008. Reflections on recent developments of the h-index and h-type indices. *Collnet Journal of Scientometrics and Information Management* 2: 1–8.

Saxena, Anurag, B.M. Gupta, and Monika Jauhari. 2011. Research performance of top engineering and technological institutes of India: A comparison of indices. *Journal of Library & Information Technology* 31: 377–381.

Schendera, Christian F.G. 2004. *Datenmanagement und Datenanalyse mit dem SAS-System*. München: Oldenbourg Wissenschaftsverlag.

Schläpfer, Frank, and Friedrich Schneider. 2010. Messung der akademischen Forschungsleistung in den Wirtschaftswissenschaften: Reputation vs. Zitierhäufigkeiten. *Perspektiven der Wirtschaftspolitik* 11: 325–339.

Schläpfer, Frank. 2012. Das Handelsblatt-BWL-Ranking und seine Zeitschriftenliste. *Bibliometrie—Praxis und Forschung* 1: Paper no. 6.

Schmitz, Christian. 2008. *Messung der Forschungsleistung in der Betriebswirtschaftslehre auf Basis der ISI-Zitationsindizes*. Köln: Eul Verlag.

Schrader, Ulf, and Thorsten Hennig-Thurau. 2009. VHB-JOURQUAL2: Method, results, and implications of the German Academic Association for Business Research's journal ranking. *BuR Business Research* 2: 180–204.

Sharma, Bharat, Sylvain Boet, Teodor Grantcharov, Eunkyung Shin, Nicholas J. Barrowman, and M. Dylan Bould. 2013. The h-index outperforms other bibliometrics in the assessment of research performance in general surgery: A province-wide study. *Surgery* 153: 493–501.

Spearman, Charles. 1904. General intelligence objectively determined and measured. *American Journal of Psychology* 15: 201–293.

Sternberg, Rolf, and Timo Litzenberger. 2005. The publication and citation output of German Faculties of Economics and Social Sciences: A comparison of faculties and disciplines based upon SSCI data. *Scientometrics* 65: 29–53.

Stock, Wolfgang G. 2001. *Publikation und Zitat: Die problematische Basis empirischer Wissenschaftsforschung*. Köln: FH Köln.

Stolz, Ingo, Darwin D. Hendel, and Aaron S. Horn. 2010. Ranking of rankings: Benchmarking twenty-five higher education ranking systems in Europe. *Higher Education* 60: 507–528.

Storbeck, Olaf. 2012. Handelsblatt BWL-Ranking 2012: Deutsche Betriebswirte fallen zurück. Handelsblatt Online. http://www.handelsblatt.com/politik/konjunktur/bwl-ranking/handelsblatt-bwl-ranking-2012-deutsche-betriebswirte-fallen-zurueck/7142160.html. Accessed 2 March 2015.

Tavenas, Francois. 2004. *Quality assurance: A reference system for indicators and evaluation procedures*. Brussels: European University Association.

Ursprung, Heinrich W. 2003. Schneewitchen im Land der Klapperschlangen: Evaluation eines Evaluators. *Perspektiven der Wirtschaftspolitik* 4: 177–189.

Ursprung, Heinrich W., and Markus Zimmer. 2007. Who is the "Platz-Hirsch" of the German Economics Profession? A citation analysis. *Jahrbücher für Nationalökonomie und Statistik* 227: 187–208.

Usher, Alex, and Massimo Savino. 2006. *A world of difference: A global survey of university league tables*. Toronto: Educational Policy Institute.

Vanclay, Jerome K. 2007. On the robustness of the h-index. *Journal of the American Society for Information Science and Technology* 58: 1547–1550.

Van Raan, A.F.J. 1996. Advanced bibliometric methods as quantitative core of peer review based evaluation and foresight exercises. *Scientometrics* 36: 397–420.

Van Raan, A.F.J. 2006. Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. *Scientometrics* 67: 491–502.

Waldkirch, Rüdiger W., Matthias Meyer, and Michael A. Zaggl. 2013. Beyond publication counts: the impact of citations and combined metrics on the performance measurement of German business researchers. *Zeitschrift für Betriebswirtschaft Special Issue* 3 (2013): 61–86.

Webster, Gregory D, Peter K. Jonason, Tatiana Orozco Schember. 2009. Hot topics and popular papers in evolutionary psychology: Analyses of title words and citation counts in Evaluation and Human Behavior, 1979–2008. *Evolutionary Psychology* 7: 348–362.

Weinstock, Melvin. 1971. Citation indexes. *Encyclopaedia of Library and Information Science* 5: 16–40.

Winkler, Oliver. 2014. Exzellente Wahl: soziale Selektivität und Handlungsorientierungen bei der Wahl von Spitzenbildung im Hochschulbereich. *Zeitschrift für Soziologie der Erziehung und Sozialisiation* 34: 280–296.

Xu, Weichao, Yunhe Hou, Y.S. Hung, and Yuexian Zou. 2013. A comparative analysis of Spearman's rho and Kendall's tau in normal and contaminated normal models. *Signal Processing* 93: 261–276.

Zhang, Chun-Ting. 2009. The e-Index: Complementing the h-Index for excess citations. *PloS One* 4: Paper e5429.

Zhu, Jia, Hassan Saeed-Ul, Hamid Turab Mirza, and Qing Xie. 2014. Measuring recent research performance for Chinese universities using bibliometric methods. *Scientometrics* 101: 429–443.

Zitt, Michel, Suzy Ramanana-Rahary, and Elise Bassecoulard. 2005. Relativity of citation performance and excellence measures: from cross-fields to cross-scale effects of field-normalisation. *Scientometrics* 63: 373–401.