Journal of Informetrics 5 (2011) 649–658

Contents lists available at ScienceDirect

Journal of Informetrics

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

# Is concentration of university research associated with better research performance?

Henk F. Moed[a,*], Félix de Moya-Anegón[b], Carmen López-Illescas[b], Martijn Visser[c]

[a] Elsevier, Radarweg 29, 1043 NX Amsterdam, The Netherlands
[b] SCImago Unit, IPP, CCHS Institute, CSIC, Spanish National Research Council, Madrid, Spain
[c] Centre for Science and Technology Studies (CWTS), Leiden University, the Netherlands

ARTICLE INFO

*Article history:*
Received 4 February 2011
Received in revised form 15 May 2011
Accepted 14 June 2011

*Keywords:*
University rankings
Citation analysis
Research performance
Research concentration
Disciplinary specialization
National academic systems
Scopus

ABSTRACT

This paper analyses relationships between university research performance and concentration of university research. Using the number of publications and their citation impact extracted from Scopus as proxies of research activity and research performance, respectively, it examines at a national level for 40 major countries the distribution of published research articles among its universities, and at an institutional level for a global set of 1500 universities the distribution of papers among 16 main subject fields.

Both at a national and an institutional level it was found that a larger publication output is associated with a higher citation impact. If one conceives the number of publications as a measure of concentration, this outcome indicates that, in university research, concentration and performance are positively related, although the underlying causal relationships are complex. But a regression analysis found *no* evidence that more concentration of research *among* a country's universities or *among* an institution's main fields is associated with better overall performance.

The study reveals a tendency that the research in a particular subject field conducted in universities specializing in other fields outperforms the work in that field in institutions specializing in that field. This outcome may reflect that it is multi-disciplinary research that is the most promising and visible at the international research front, and that this type of research tends to develop better in universities specializing in a particular domain and expanding their capabilities in that domain towards other fields.



## 1. Introduction

In many OECD countries ministers responsible for higher education institutions and scientific research established a research policy aimed at concentrating research in centers of excellence, or they started distributing parts of the research budget among research institutions on the basis of performance criteria (OECD, 2010; Hicks, 2010). The latter policy is believed to be one of the key strategies to establish concentration among national research institutions, at least at a long run. What are the effects of concentration policies upon the overall research performance of a national research system and upon individual research institutions?

The optimal distribution of research funding and activity does not only relate *at a national level* to the question as to how to distribute funds *among institutions*, but also *within an institution* to distribute funds *among the various subject fields*.

* Corresponding author. Tel.: +31 204853436; fax: +31 204853354.
*E-mail addresses:* h.moed@elsevier.com (H.F. Moed), felix.moya@scimago.es (F. de Moya-Anegón), carmenlillescas@cchs.csic.es (C. López-Illescas), visser@cwts.leidenuniv.nl (M. Visser).

1751-1577/$ – see front matter 
doi:10.1016/j.joi.2011.06.003

Research institutions, especially universities, differ as regards their disciplinary specialization; while some are general, others are specialized and focus on a few main fields only. The relationship between size, specialization and performance in scientific and technological activity is a key issue in policy debates. While some are in favor of establishing more integration and breadth, others underline the need for specialization, diversity and competition (Von Tunzelmann, Ranga, Martin, & Geuna, 2003). A series of studies on the scientific and technological performance of research institutions, firms and countries analyzed these relationships. For a review the reader is referred to Brusoni and Geuna (2004).

This paper focuses on one particular type of institutions: higher education institutions, also denoted as universities throughout this paper. University rankings have gained a strong interest both from managers, researchers and the general public (e.g., CEPES, 2006; Liu & Cheng, 2005; Liu & Cheng, 2008; Salmi, 2009; SJTU, 2007; Van Raan, 2005). Although such rankings are *marketing* tools rather than *research management* tools (AUBR, 2010), underlying data constitute a rich source for secondary analyses of policy-relevant issues that help testing policy assumptions and interpreting rankings in a proper way (e.g., Calero-Medina, López-Illescas, Visser, & Moed, 2008; López-Illescas, de Moya-Anegón, & Moed, 2011). From this perspective, in a quantitative-empirical, bibliometric approach, using citation impact as proxi of research performance and the number of published articles as an indicator of research activity, this paper analyses 40 major national university systems and 1500 universities, addressing the following four research questions:

1. Do countries in which research is concentrated in a relatively small number of universities perform better than countries do that show a more even distribution of research between their higher education institutions?
2. Do countries in which universities tend to show a large degree of overall disciplinary specialization perform better than nations do with more general universities covering a wide range of subject fields?
3. Do specialized universities perform better than general academic institutions do?
4. Do universities specializing in a particular subject field outperform in their field of specialization universities that are less specialized in that field?

The relation between ‘size’ and ‘performance’ is often analyzed in terms of critical mass, i.e., a minimum size of research input that is needed to create substantial scientific progress. Critical mass is to be assessed at the level of a research group or department rather than at the level of a university as a whole. If a university published, say, one hundred papers per year in a particular research subject, a key question is whether this output was created by one single research department carrying out a coherent research programme, or whether it was dispersed among a large number of small groups or individuals hardly interacting with each other. The data analyzed in this paper relate to a breakdown of a university’s research output into 16 main subject fields; data at the level of research groups or departments were *not* available in the current study.

Concentration of research can be considered from *two* viewpoints: *internal* and *external*. The first compares the publication output of a university in a particular field to the *same* institution’s output in other fields, and to its total output; the second compares it to the number of articles published by *other* institutions in the same subject field. A ‘big’ research university may show *internally* a low publication activity in a field compared to its output in other fields, but *externally*, compared to other institutions in the same field, be among the most productive ones. A similar distinction can be made in the analysis of national university systems: one can apply a internal, national viewpoint, comparing one university with other universities in the *same* country, or an external, international one, comparing it with all *other* institutions in the world. Therefore, when interpreted from an external viewpoint, the number of publications is *also* an indicator of *concentration* of research.

The structure of this paper is as follows. Section 2 gives a description of the data analyzed and methodologies applied. Results at the level of national university systems are presented in Section 3, and those conducted at the level of individual universities in Section 4. Finally, Section 5 discusses the outcomes and makes suggestions for further research.

## 2. Data and methodology

This paper presents a secondary analysis of bibliometric data on research institutions extracted from the SCImago Institutions Rankings (SIR) Database (SCImago, 2010), which is based on *Scopus*. Scopus is a comprehensive, multi-disciplinary citation index published by Elsevier. It covers nearly 18,000 titles including 16,500 peer-reviewed journals (Scopus, 2010). SIR presents bibliometric indicators of the nearly 3000 most productive research institutions across the globe in terms of number of published articles. For the current paper a dataset was created with bibliometric indicators for the 1500 higher education institutions with the highest number of research articles published during the time period 2003–2007. The dataset contains for each of these 1500 universities and for each of 27 scientific main fields the number of published papers per year during 2003–2007, and the average number of citations per paper received up until 2007.

The indicators calculated in this study are described in detail in Table 1. The study applies a categorization of journals into 16 main fields, created by condensing the 27 categories available in Scopus and the SCImago database by combining a number of small fields with their most closely related big field. A full list of main fields is given in Table 4 in Section 4.

For each institution an *overall disciplinary specialization index* (ODS) is calculated. It measures the extent to which a university’s papers are evenly distributed across main subject fields, taking into account the world distribution of papers across fields. It is defined as Gini’s Index for a university’s distribution of the *publication activity indices (PAI)* across disciplines. A value of 0 is obtained for a university for which the distribution of published papers across fields is equal to the world distribution. More details can be found in Egghe and Rousseau (1990) and Rousseau (1992). The degree of ‘overall’ specialization

**Table 1**
Bibliometric indicators calculated in this paper.

| Indicator | What it measures | Definition |
|---|---|---|
| University Viewpoint | | |
| Number of publications | Research activity | Let $p_{ij}$ be the number of articles published by university $i$ in field $j$, $i = 1,\ldots,n$, $j = 1,\ldots,m$, and let $c_{ij}$ be the total number of citations received by these $p_{ij}$ publications. The total number of articles published by university $i$ is given by $P_i = \sum_j p_{ij}$ |
| Publication activity index (PAI) | The extent to which a university focuses its publication activity on a particular research field; in other words, it indicates the degree of a university's specialization in a particular main field. | The publication activity index $A_{ij}$ of university $i$ in field $j$ is defined as $A_{ij} = \frac{p_{ij}/\sum_j p_{ij}}{\sum_i p_{ij}/\sum_i\sum_j p_{ij}}$. The normalized publication activity index $A_{ij}^N$ of university $i$ in field $j$ is defined as $A_{ij}^N = \frac{A_{ij}-1}{A_{ij}+1}$. |
| Overall disciplinary specialisation index (ODS) | The extent to which a university's research papers are evenly distributed among research fields (as in general universities) or are more concentrated in particular fields (as for instance in medical, agricultural or certain technical universities) | The overall disciplinary specialization of university $I$, $S_i$, is defined as the Gini index of the distribution $\left\{\frac{A_{i1}}{\sum_j A_{ij}}, \frac{A_{i2}}{\sum_j A_{ij}}, \ldots, \frac{A_{im}}{\sum_j A_{ij}}\right\}$. |
| Field-normalised citation impact per paper, also labelled shortly as citation impact | Intellectual influence; prominence of research groups in their fields; their authoritativeness; their visibility at the international research front | The Relative Citation Rate $r_{ij}$ of the articles published by university $i$ in field $j$ is defined as $r_{ij} = \frac{p_{ij}/c_{ij}}{\frac{1}{n}\sum_i p_{ij}/c_{ij}}$ In Section 4, $r_{ij}$ is based on a 'national' average (see main text in Section 2). The overall relative citation rate $R_i$ of the articles published by university $i$ is defined as $R_i = \frac{\sum_j p_{ij}.r_{ij}}{\sum_j p_{ij}}$ |
| National Viewpoint | | |
| Degree of concentration among a country's universities | The degree of concentration of published research articles among universities in a national university system | If we assume that a specific country $L$ has $k$ universities, and if $P^L$ indicates the number of articles published by university $i$ in $L$, $i = 1,\ldots,k$, the degree of concentration of article output among $L$'s universities is defined as the Gini index for the distribution $\{P^L, P^L, \ldots, P^L\}$ |
| The mean and standard deviation of the number of publications over a country's universities | The mean of number of publications of a country's universities and its variability across universities | The mean $\overline{P}^L$ of the article output of $L$'s universities is defined as $\overline{P^L} = \frac{1}{k}\sum_i P_i^L$. The variability in article output among universities in L is defined as the Standard Deviation (SD) of the distribution. |
| The mean and standard deviation of the citation impact over a country's universities | The mean of the field-normalised citation impact per paper over a country's universities, and its variability across universities | If $R^L$ indicates the Overall Relative Citation Rate of university $i$ in L, $\overline{R}^L$, the Mean Overall Disciplinary Specialization is defined as $\overline{R^L} = \frac{1}{k}\sum_i R_i^L$ The variability among universities in L is defined as the Standard Deviation (SD) of the distribution. |
| The mean and standard deviation of the overall disciplinary specialization over a country's universities | The mean degree of disciplinary specialization within a country's universities and the variability across universities | If $S^L$ indicates the Overall Disciplinary Specialization in university $i$ in L, $\overline{S}^L$, the Mean Overall Disciplinary Specialization is defined as $\overline{S^L} = \frac{1}{k}\sum_i S_i^L$. The variability among universities in L is defined as the Standard Deviation (SD) of the distribution. |

of a university as a whole and its specialization in a particular field are dependent. General institutions will hardly show a very strong specialization in a particular field. Plotting for each university its overall disciplinary concentration (on the vertical axis) against its PAI in a particular field results in a V-shaped scatter plot, indicating that in the subset of institutions with extreme PAI values, i.e., with high _or_ low publication activity indices, specialized universities are overrepresented.

In Section 3, analyzing national university systems from a global perspective, the citation impact indicators are normalized, relative to '*world*' average based on *all* 1500 universities analyzed. But in the institutional analysis in Section 4, in order to reduce biases due to differences between countries in the state of scientific development and international visibility between nations, an institution's normalized citation impact in a particular field is calculated by dividing its citation-per-article ratio in a field by the average citation rate in that field calculated over all universities *in the same country*, thus comparing for instance the citation impact of a US university with a national US average, and that of a Chinese higher education institution with a an average for China.

This study aims at analyzing the relationship of *both* dimensions of concentration distinguished in the Introduction (internal and external) with performance. Therefore, in the statistical analysis a multiple regression analysis is carried out with indicators of both types of concentration of publications as independent variables.

## 3. Results for national university systems

Fig. 1 characterizes 40 major national university systems, presenting the variation in their universities' number of publications (on the *x*-axis) and their citation impact (*y*-axis). China, Russia and Japan show the largest variation in article output, and again China and Russia, and also Mexico, South Africa and Chile the largest variability in citation impact. Switzerland, Belgium and Netherlands reveal the lowest variability in citation impact across their universities. USA, UK, Australia and Canada are at an intermediate level.

Fig. 2 shows that the statistical relationship between a country's degree of concentration of research publications among universities on the one hand, and mean citation impact over its universities on the other, is complex. According to Table 2 Pearson's $R$ is $-0.39$ and Spearman's Rho $-0.44$ (both significant at $p = 0.01$). The top league in terms of citation impact reveals an Anglo-Saxon branch of countries showing a strong concentration (including USA, UK, Australia, Canada), and a Western-European group with nations in which research is more evenly distributed among their universities, including Switzerland, Netherlands, Belgium, Denmark and Sweden. Interestingly, the Asian countries Japan and Taiwan seem to follow the Anglo-Saxon, and Hong Kong the European model. The latter also applies to Israel.

Fig. 3 shows that the mean degree of overall disciplinary specialization (ODS) of a nation's universities shows a rather strong negative correlation to the mean citation impact over its universities ($n = 40$, Pearson's $R = -0.71$, $p = 0.001$; Spearman's $R = -0.73$, $p = 0.001$). Belgium shows the lowest mean of ODS among its universities, and China and Russia the highest. Almost half of countries have values between 0.4 and 0.5; all countries with a citation impact above 1.5 except Belgium are in this range. Despite the observed correlation that is further interpreted in Section 5, this figure clearly shows the wide range of citation impact values for one and the same degree of ODS. Countries such as USA, the Netherlands or Sweden with impact

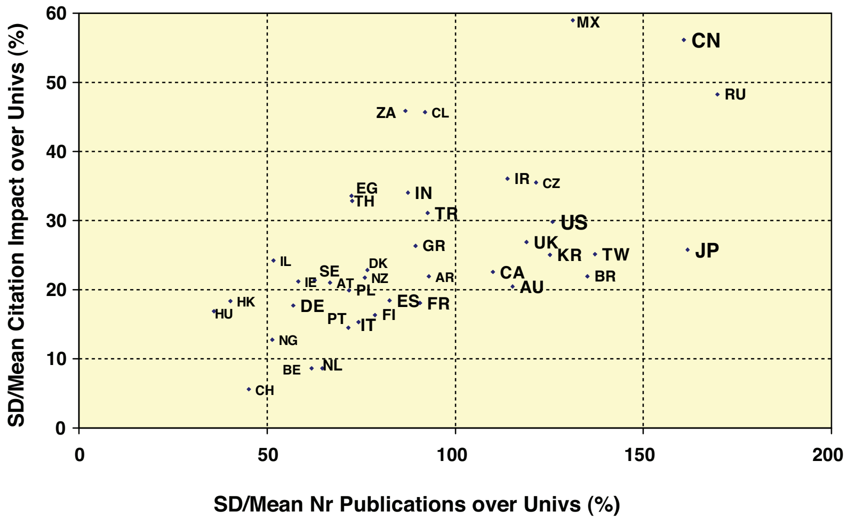


**Fig. 1.** Variation in universities' number of publications and their citation impact in 40 countries. Figures show results for the 40 countries containing 6 or more universities in the set of 1500 higher education institutions analyzed in this paper. SD means Standard Deviation. For the precise definition of the variables plotted in the figures the reader is referred to Table 1. Font size reflects number of universities hosted: super large: >100 universities; large: 35–100; medium: 9–34; small: 6–8 universities. For country abbreviations (alpha 2 codes) see http://en.wikipedia.org/wiki/ISO_3166-1.

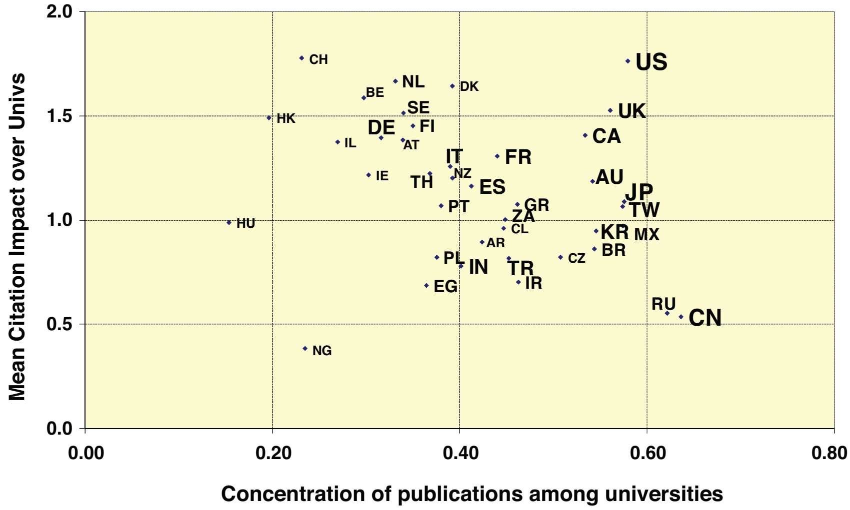


**Fig. 2.** Mean citation impact versus degree of concentration of published papers among universities for 40 major countries. Figures show results for the 40 countries containing 6 or more universities in the set of 1500 higher education institutions analyzed in this paper. SD means Standard Deviation. For the precise definition of the variables plotted in the figures the reader is referred to Table 1. Font size reflects number of universities hosted: super large: >100 universities; large: 35–100; medium: 9–34; small: 6–8 universities. For country abbreviations (alpha 2 codes) see http://en.wikipedia.org/wiki/ISO_3166-1.

values above 1.5 have ODS values that are statistically similar to that of countries such as Chile, Brazil or Egypt which show impact values below world average.

Table 2 shows that a country's mean citation impact over its universities correlates strongly with the average number of published papers per institution. However, when a multiple regression model is applied, with the mean citation impact over a country's universities as the dependent variable, and the mean number of publications, mean ODS, and the concentration of papers among universities as independent variables, the statistical effect of the latter variable is not significant.

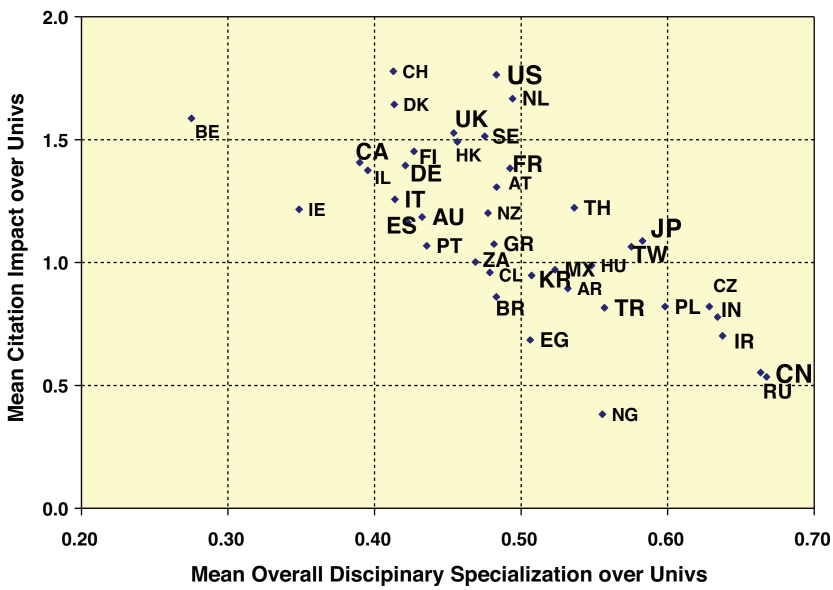


**Fig. 3.** Mean citation impact versus mean of overall disciplinary specialization over universities in 40 major countries. Figures show results for the 40 countries containing 6 or more universities in the set of 1500 higher education institutions analyzed in this paper. SD means Standard Deviation. For the precise definition of the variables plotted in the figures the reader is referred to Table 1. Font size reflects number of universities hosted: super large: >100 universities; large: 35–100; medium: 9–34; small: 6–8 universities. For country abbreviations (alpha 2 codes) see http://en.wikipedia.org/wiki/ISO_3166-1.

**Table 2**
Linear correlation and regression analysis for countries.

Correlation analysis

| | Mean no. of publications over universities | | Mean overall disciplinary concentration over universities | | Concentration of papers among universities | |
|---|---|---|---|---|---|---|
| | *R* | Rho | *R* | Rho | *R* | Rho |
| Mean citation impact over universities | 0.83* | 0.85* | −0.71* | −0.73* | −0.39* | −0.44* |
| Mean no. of publications over universities | | | −0.61* | −0.66* | −0.29 | −0.25 |
| Mean overall disciplinary concentration over universities | | | | | 0.40* | 0.41* |

Regression analysis

| Dependent variable | Independent variables | | | | | |
|---|---|---|---|---|---|---|
| | Mean no. of publications over universities | | Mean overall disciplinary concentration over universities | | Concentration of papers among universities | |
| | Beta × 1000 | Sign | Beta | Sign | Beta | Sign |
| Mean citation impact over universities | 0.079 | * | −1.11 | * | −0.272 | n.s. |

If MODS denotes the mean of overall disciplinary specialization over a country's universities, MP the mean of total number of publications over universities, MCI the mean citation impact over universities and CPU the degree of concentration of papers among a country's universities, the model applied in the multiple regression analysis is: MCI = $\beta 1 \times$ MODS + $\beta 2 \times$ MP + $\beta 3 \times$ CPU. The number of countries involved is 40. The *R*-square of the model is 0.76 ($p < 0.01$). An asterisk indicates whether a value deviates significantly from zero at $p < 0.01$. 'n.s.' means 'not significant at $p = 0.01$'. *R*: Pearson correlation coefficient; Rho: Spearman rank correlation coefficient.

## 4. Results at the level of individual universities

It needs emphasizing that in the analyses presented in this section a university's citation impact is calculated relative to a *national* average. This is justified as the purpose of this chapter to analyze differences in overall disciplinary specialization (ODS) of universities, rather than assessing their research performance from a global perspective.

Analyzing a university as a whole, and calculating indicators related to *all* its publications, the upper half of Table 3 reveals that universities showing a high citation impact tend to have a larger publication output (P) and a lower ODS than universities have that generate a low citation impact. However, a linear, multiple regression analysis presented in the lower part of Table 3, – defining citation impact as dependent variable, and ODS and P as independent variables –, shows that the unique effect of ODS upon citation impact is not statistically significant.

In other words, when restricting the attention to the subsets of the data that have a similar value for the number of publications, there is no significant difference in citation impact between general and specialized universities. It must be noted that the predictor variables are correlated with each other; according to the upper half of Table 3 the Pearson's *R* calculated for number of publications and overall disciplinary specialization amounts to −0.54.

Table 4 presents the outcomes of a multiple regression analysis at the level of *main fields* within a university. In order to eliminate extremely high values of the dependent variable, universities publishing less than 10 papers in a main field during 2003–2007 were eliminated from the analysis of that main field, thus reducing the total number of cases studied with 14

**Table 3**
Linear correlation and regression analysis for universities.

Correlation analysis

| | Total publications | | Overall disciplinary specialization | |
|---|---|---|---|---|
| | *R* | Rho | *R* | Rho |
| Citation impact | +0.31 * | +0.50 * | −0.14 * | −0.25 * |
| Total publications | | | −0.40 * | −0.54 * |

Regression analysis

| Dependent variable | Independent variables | | | |
|---|---|---|---|---|
| | Total publications | | Overall disciplinary specialization | |
| Citation impact | Beta × 1000 | Sign | Beta | Sign |
| | 0.0175 | * | −0.050 | n.s. |

If ODS denotes overall disciplinary specialization, *P* the total number of publications and CI the citation impact, the model applied in the multiple regression analysis was: CI = $\beta 1 \times$ ODS + $\beta 2 \times P$. The *R*-square of the model is 0.10 ($p < 0.01$). An asterisk indicates whether a value deviates significantly from zero at $p < 0.01$. 'n.s.' means 'not significant'. *R*: Pearson correlation coefficient; Rho: Spearman rank correlation coefficient.

**Table 4**
Multiple regression coefficients per main field.

| Main field | Overall disciplinary specialization (ODS) | | Publication activity index (PAI) | | Number of publications | | Model | |
|---|---|---|---|---|---|---|---|---|
| | β1 | Sign | β2 | Sign | β3 × 1000 | Sign | *R*-square | ModSign |
| Agriculture & Biological Sciences | +0.078 | n.s. | −0.385 | * | +0.11 | * | 0.079 | * |
| Arts & Humanities | −0.900 | n.s. | −0.055 | n.s. | +1.39 | n.s. | 0.024 | * |
| Biochemistry & Molecular Biology | +0.076 | n.s. | −0.065 | n.s. | +0.11 | * | 0.073 | * |
| Chemistry | +0.065 | n.s. | −0.058 | n.s. | +0.29 | * | 0.104 | * |
| Computer Science | −0.067 | n.s. | −0.438 | * | +0.32 | * | 0.066 | * |
| Earth & Planetary Sciences | +0.172 | n.s. | +0.106 | n.s. | +0.36 | * | 0.043 | * |
| Economics and Business | −0.011 | n.s. | −0.263 | * | +0.41 | * | 0.030 | * |
| Engineering | −0.255 | * | −0.627 | * | +0.06 | * | 0.163 | * |
| Environmental Sciences | −0.263 | * | −0.156 | * | +0.34 | * | 0.044 | * |
| Health Professions and Nursing | +0.384 | n.s. | −0.597 | * | +0.72 | * | 0.073 | * |
| Materials science | +0.049 | n.s. | −0.530 | * | +0.20 | * | 0.099 | * |
| Mathematics | +0.141 | n.s. | −0.124 | n.s. | +0.40 | * | 0.039 | * |
| Medicine | −0.031 | n.s. | −0.245 | * | +0.06 | * | 0.056 | * |
| Other Life Sciences | −0.099 | n.s. | −0.103 | * | +0.09 | * | 0.035 | * |
| Physics and Astronomy | +0.054 | n.s. | −0.303 | * | +0.13 | * | 0.046 | * |
| Social Sciences and Psychology | +0.292 | n.s. | −0.301 | * | +0.28 | * | 0.020 | * |

If ODS denotes the overall disciplinary specialization, PAI the publication activity, *P* the number of publications, and CI their citation impact, the model applied in the analysis is: CI = β1 × ODS + β2 × PAI + β3 × *P*. Column sign: an asterisk indicates that a value of beta deviates significantly from zero at $p < 0.01$. 'n.s.' means 'not significant'. ModSign: an asterisk indicates whether the model is significant at $p < 0.01$.

per cent. The model applied is specified in the legend to Table 4. The model is significant for all main fields, although the explained variance is low, ranging between 0.02 and 0.16.

The analysis by main field of the statistical effect of ODS and number of publications upon citation impact are consistent with the observations for universities as a whole. In all fields except two, ODS does not have a statistically significant effect upon citation impact. The exceptions are *Engineering* and *Environmental Sciences*, where it is negative. Next, in all fields except *Arts & Humanities* the number of publications has a statistically positive effect upon citation impact.

Interestingly, in all fields except *Earth & Planetary Sciences*, publication activity index has a ***negative*** statistical effect upon citation impact. For 11 fields the effect is statistically significant. Exceptions are five basic-oriented fields: *Arts & Humanities*, *Biochemistry & Molecular Biology*, *Chemistry*, *Mathematics* and *Earth & Planetary Sciences*. The lowest values of the regression coefficient can be found in the applied fields *Engineering*, *Health Professions & Nursing*, *Materials Science* and *Computer Science.*

As an illustration of differences between main fields, Fig. 4 gives a scatter plot of a university's citation impact versus its normalized publication activity index (PAI) in two specific fields: *Engineering* and *Biochemistry & Molecular Biology*. As

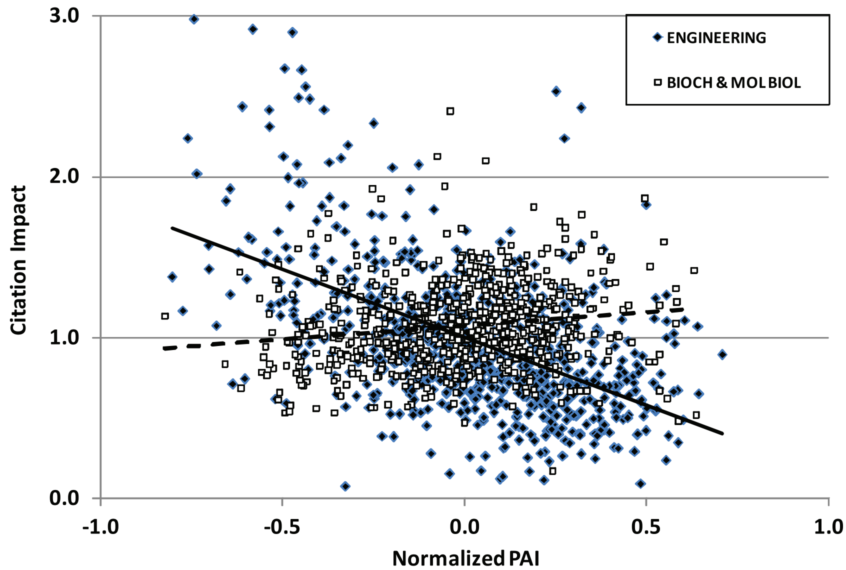


**Fig. 4.** Citation impact versus PAI for universities in Engineering and in Biochemistry & Molecular Biology. PAI: Publication Activity Index. For the definitions of normalized PAI and citation Impact see Table 1 in Section 2. Data relate to universities publishing at least 10 papers per year in a field, and located in Australia, Canada, China, Germany, France, Great Britain, Japan, South Korea and USA. The number of universities amounts to 740 in Engineering and 644 in Biochemistry & Molecular Biology. The straight lines represent the linear regression lines based on the least squares principle.

**Table 5**
Fields of specialization of universities with a low PAI but high citation impact in a particular field.

| | Agr | A&H | Bioch | Chem | Comp | Earth | Econ | Eng | Envir | Health | Mater | Math | Med | Life | Phys | Soc Sc |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Agr | | | | | | | | | | 8 | | | 7 | | | |
| A&H | | | | | | | | | | | | | | | | |
| Bioch | | | | | | | | | | | | | | | | |
| Chem | | | | | 3 | | | | | 3 | | | 3 | | | |
| Comp | | | | | | | | | | 3 | | | 4 | 3 | | |
| Earth | | | | | | | 3 | | | | | | | | | |
| Econ | | | | | | | | | | | | | | | | |
| Eng | 6 | | 5 | 4 | | | | | 3 | 12 | | | 15 | 7 | | |
| Envir | | | | | | | | | | | | | | | | |
| Health | | | | | 3 | | | 5 | | | 4 | | | | | |
| Mater | 4 | | | | | | | | | 8 | | | 9 | 3 | | |
| Math | | | | | | | | | | | | | | | | |
| Med | 5 | 4 | | | 4 | 7 | 4 | 3 | 7 | | 3 | 4 | | | 3 | |
| Life | | | | | | | 3 | 4 | | | 6 | | | | | |
| Phys | | | | | | | | | | 7 | | | 12 | 4 | | |
| Soc Sc | | | | | | | | | | | | | | | | |

The table presents for each main field and for institutions with a normalized PAI below −0.5 and a relative citation impact above 1.5 in that field (in the rows) the number of universities with a normalized PAI above +0.5 in the various fields (in the columns). For instance, in Engineering, 15 institutions showing a low PAI but high impact specialize in medicine. The table shows data only for cells in which the number of universities is at least 3.

outlined in Section 2, universities showing a very low normalized PAI (e.g., below −0.5) in one field tend to specialize in other main fields.

Table 5 sheds more light upon the negative statistical effect of PAI upon citation impact. It gives for each main field, and for universities showing a strong under-activity but at the same time a high citation impact in that field, an overview of the main fields in which these universities are specialized. The majority of universities with a low PAI but a high citation impact in *Engineering*, *Materials Science* or *Computer Science*, are specializing in *Medicine* and *Health Professions & Nursing*. Noteworthy is also the strong relationship between *Physics* and *Medicine*. On the other hand, institutions with a low PAI but a high citation impact in *Health Professions & Nursing* show an over-activity in the three applied fields mentioned above, while institutions with under-activity but high citation impact in *Medicine* tend to be highly active especially in *Earth Sciences* and *Environmental Science* and *Agriculture & Biological Sciences*.

## 5. Discussion and conclusions

This paper analyzed statistical relationships between research performance and two types of research concentration: one related at the level of a country to the distribution of published research articles among its universities, and a second at the level of an individual university to the distribution of papers among research fields. Analyzing in Scopus a global set of 1500 universities, and using citation impact as performance indicator, the overarching research question was: Does more concentration lead to better research? The general conclusion of four separate analyses presented in this study is as follows: if one adopts an external viewpoint, conceiving the volume of publication output as an indicator of concentration, performance and concentration are statistically associated; but analyzing concentration from an internal viewpoint, i.e., within a country among its universities or within a university among subject fields, they are – at least according to three analyses – *not* correlated in a statistically significant manner.

A *first* analysis found in a group of 40 major *countries* that a country's mean citation impact over its universities strongly correlates with the average number of published papers per institution. If one adopts an external viewpoint and conceives the publication output as a measure of concentration, this result shows that concentration and performance are associated. But the underlying causal relationships are complex: more 'input' may lead to more papers and a higher citation impact, but in a system in which research funding is at least partly performance-based, it can be expected to be the other way around as well.

The analysis also revealed a negative, though weak linear or rank correlation between a country's mean citation impact over all its universities and the degree of concentration of research among its universities. However, in a multiple regression analysis, the unique effect of the latter upon the former is not statistically significant. Among the top performing nations one finds both concentrated and more evenly distributed national university systems. One should note the effect of China on the negative correlation between these two variables. The low average citation impacts of scientifically developing countries are also affected by other factors such as the so called database coverage effect. That is, the recent inclusion of numerous more nationally oriented journals in the international databases Scopus or Web of Science tend to provoke the paradoxical effect that countries profiting most of the database expansion in terms of percentage of published documents tend to show a *decline* in their average citation rate (López-Illescas, de Moya-Anegón, & Moed, 2009). This factor needs to be taken into account when interpreting the correlation between the two variables.

Although in a *second* analysis the average degree of disciplinary specialization of a nation's universities did show a significant, negative linear correlation with their average citation impact, confounding factors make the interpretation of this finding difficult. The underlying causal relations are complex, and other variables have to be taken into account as well. The above mentioned 'database coverage effect' which lowers the average impacts of countries such as China plays an important role here too. The results presented in this paper enable one to examine one of such factors in more detail: a university's degree of disciplinary specialization. This analysis corrects for differences in a country's state of scientific development and in database coverage between countries by calculating indicators that assess a university's citation impact in a research field relative to a *national* average.

At the level of *individual universities,* a *third* analysis revealed that universities with a large publication output tend to have a higher citation impact than institutions have that publish lower numbers of articles. This is also true in all but one main fields. If one conceives the number of published papers as an indicator of concentration, this outcome shows that more concentration is associated with higher citation impact. But, as noted in the paragraph on national university systems, the underlying causal relationships between 'input' and 'output' are complex.

It was also found that universities showing a high overall disciplinary specialization (ODS) tend to have a lower citation impact than general universities do. But the linear or rank correlation is weak, and there are many specialized universities with an outstanding performance. A linear multiple regression analysis revealed that the unique effect of ODS upon citation impact is not statistically significant. Hence, from an internal, institutional viewpoint there is no evidence that more concentration among main fields within a university is associated with better overall performance. A similar conclusion can be drawn within all subject fields except two: *Engineering* and *Environmental Sciences*.

But a *fourth* analysis indicated a perhaps counter-intuitive tendency that when universities show a strong *under*-activity in a subject field, their papers in that field tend to have a higher citation impact than other articles in that field have that were published from institutions showing a higher publication activity therein. Since universities showing an under-activity in a field must be specializing in other fields, the outcomes reveal in most fields a weak tendency that the research in a particular subject field conducted in universities specializing in other fields outperforms the work in that field in institutions specializing in that field. An additional analysis revealed that the institutions with a strong under-activity but at the same time a high citation impact in *Engineering*, *Materials Science* and *Computer Science*, are themselves specializing in (bio-)medical research, whereas institutions with a low PAI but high impact in (bio-)medical and health related fields tend to be focusing not only on particular applied, but also on basic science fields.

These results may reflect that it is multi-disciplinary research that is the most promising and visible at the international research front, and that this type of research tends to develop better in universities specializing in a particular domain and expanding their capabilities in that domain towards other fields. If specialization is too strong, an institution may not be able to pick up the developments in emerging topics that require a structural contribution from fields in which it hardly shows activity and does not have expertise.

More research is needed at the level of research groups and their embedment within their institution to test these hypotheses. The further development and application of tools to identify and measure multi- or interdisciplinary research networks within and across universities plays a key role in this research. Differences among research fields should be studied in more detail and look for explanations why some fields behave differently, for instance, by investigating whether their behavior is uniform across national university systems.

As regards applied sciences and engineering it must be noted that the performance measures applied in this paper do *not* measure the contribution to economic or technological progress in a direct way, but rather its reflection in the international scientific literature. To the extent that technological performance leaves its traces in the scientific literature at all, the outcomes suggest that science and technology are becoming closer and closer, that major technological achievements are more and more science based; research in the 'Pascal quadrant' becomes increasingly important and tends to develop better in universities specializing in more basic-oriented fields than in highly specialized technological institutions.

Finally, It needs to be underlined that the results presented in this paper refer only to the *higher education* sector, and cannot be directly extended to other research institutions or sectors. In fact, a preliminary analysis of all 3000 especially in the fields where the overwhelming part of the top research is carried out in those specialized research centers. Therefore, a further interpretation of the results obtained in this paper should also take into account the structure of a country's national academic system and the organisation of its research, specially the research centres founded outside the university system. In order to generate a more complete overview, a follow-up study should examine the relationship between research concentration and performance in all types of research institutions and institutional sectors.

## References


AUBR. (2010). Expert Group on the Assessment of University-Based Research. Assessing Europe's University-Based Research. European Commission – DG Research. Retrieved 1 February 2011 from http://ec.europa.eu/research/era/docs/en/areas-of-actions-universities-assessing-europe-university-based-research-2010-en.pdf

Brusoni, S., & Geuna, A. (2004). Specialisation and integration. In H.F. Moed, W. Glänzel, U. Schmoch (Eds.), *Handbook of quantitative science and technology research. The use of publication and patent statistics in studies of S&T systems* (pp. 733–758). Dordrecht (the Netherlands): Kluwer Academic Publishers.

Calero-Medina, C., López-Illescas, C., Visser, M. S., & Moed, H. F. (2008). Important factors in the interpretation of bibliometric rankings of world universities. *Research Evaluation*, *17*, 71–81.

CEPES. (2006). The Berlin Principles on Ranking of Higher Education Institutions. Retrieved January 11, 2009, from http://www.cepes.ro/themes/policy_reform/quality_assurance/ranking/Berlin%20Principles.pdf

Egghe, L., & Rousseau, R. (1990). *Introduction to informetrics*. Amsterdam: Elsevier.

Hicks, D. (2010). Overview of models of performance-based research funding systems. In *Performance-based funding for public research in tertiary education institutions: workshop proceedings*. OECD Publishing. http://dx.doi.org/10.1787/9789264094611-en

Liu, N. C., & Cheng, Y. (2005). The Academic Ranking of world universities – Methodologies and problems. *Higher Education in Europe*, *30*(2), 127–136.

Liu, N. C., & Cheng, Y. (2008). Examining major rankings according to the Berlin principles. *Higher Education in Europe*, *33*(2/3), 201–208.

López-Illescas, C., de Moya-Anegón, F., & Moed, H. F. (2009). Comparing bibliometric country-by-country rankings derived from the Web of Science and Scopus: The effect of poorly cited journals in oncology. *Journal of Information Science*, *35*, 244–256.

López-Illescas, C., de Moya-Anegón, F., & Moed, H. F. (2011). A ranking of universities should account for differences in their disciplinary specialization, submitted for publication.

OECD. (2010). *Performance-based funding for public research in tertiary education institutions: workshop proceedings*. OECD Publishing. http://dx.doi.org/10.1787/9789264094611-en

Rousseau, R. (1992). Specialization and diversity in informetric research. Ph.D. Thesis, University of Antwerp.

Salmi, J. (2009). *The challenge of establishing world-class universities*. Washington, D.C., USA: World Bank.

SCImago. (2010). The SIR 2009 world report. Retrieved 1 February 2011 from http://www.scimagoir.com/. Last accessed 1 February 2011

Scopus. (2010). What does it cover? Retrieved 1 February 2011 from http://info.scopus.com/scopus-in-detail/facts/. Last accessed on 8 august 2010.

SJTU. (2007). Academic ranking of world universities. Shanghai Jiao Tong University, Retrieved 22 July 2009, from http://www.arwu.org/rank/2007/ranking2007.htm

Van Raan, A. F. J. (2005). Fatal attraction – Conceptual and methodological problems in the ranking of universities by bibliometric methods. *Scientometrics*, *62*(1), 133–143.

Von Tunzelmann, N., Ranga, M., Martin, B., & Geuna, A. (2003). The effects of size on research performance – A SPRU review. Report prepared for the Office of Science and Technology, Department of Trade and Industry. Last retrieved on 1 September 2010 at www.sussex.ac.uk/Users/prff0/Publications/Size%20HE%20research.pdf