



Universiteit  
Leiden  
The Netherlands

## **Self-citations at the meso and individual levels: effects of different calculation methods**

Costas Comesana, R.; Leeuwen, T.N. van; Bordons, M.

### **Citation**

Costas Comesana, R., Leeuwen, T. N. van, & Bordons, M. (2010).  
Self-citations at the meso and individual levels: effects of different  
calculation methods. *Scientometrics*, 82(3), 517-537.  
doi:10.1007/s11192-010-0187-7

Version: Publisher's Version

License: [Creative Commons CC BY-NC 4.0 license](#)

Downloaded from: <https://hdl.handle.net/1887/78135>

**Note:** To cite this publication please use the final published version  
(if applicable).

## Self-citations at the meso and individual levels: effects of different calculation methods

Rodrigo Costas · Thed N. van Leeuwen · María Bordons

Received: 11 May 2009 / Published online: 17 February 2010

© The Author(s) 2010. This article is published with open access at Springerlink.com

**Abstract** This paper focuses on the study of self-citations at the meso and micro (individual) levels, on the basis of an analysis of the production (1994–2004) of individual researchers working at the Spanish CSIC in the areas of Biology and Biomedicine and Material Sciences. Two different types of self-citations are described: author self-citations (citations received from the author him/herself) and co-author self-citations (citations received from the researchers' co-authors but without his/her participation). Self-citations do not play a decisive role in the high citation scores of documents either at the individual or at the meso level, which are mainly due to external citations. At micro-level, the percentage of self-citations does not change by professional rank or age, but differences in the relative weight of author and co-author self-citations have been found. The percentage of co-author self-citations tends to decrease with age and professional rank while the percentage of author self-citations shows the opposite trend. Suppressing author self-citations from citation counts to prevent overblown self-citation practices may result in a higher reduction of citation numbers of old scientists and, particularly, of those in the highest categories. Author and co-author self-citations provide valuable information on the scientific communication process, but external citations are the most relevant for evaluative purposes. As a final recommendation, studies considering self-citations at the individual level should make clear whether author or total self-citations are used as these can affect researchers differently.

**Keywords** Self-citations · Micro-level · Meso-level · Individual scientists · Bibliometric indicators · Citation analysis

---

R. Costas (✉) · M. Bordons

Instituto de Estudios Documentales sobre Ciencia y Tecnología (IEDCYT), Centre for Human and Social Sciences (CCHS), Spanish National Research Council (CSIC), Madrid, Spain  
e-mail: rcostas@cwts.leidenuniv.nl

T. N. van Leeuwen

Centre for Science and Technology Studies (CWTS), Leiden University, Leiden, The Netherlands

## Introduction

In as far as Information Science is concerned, self-citations are deemed a natural, regular and indispensable part of scientific communication, since they reflect the continuous and cumulative nature of the research process (Pichappan and Sarasvady 2002). However, in evaluative bibliometrics, self-citations are often regarded as distortions that affect the validity of citations as measures of scientific impact (Schubert et al. 2006) on the grounds that they do not reveal anything about the impact of a work beyond their own producers. In this sense, self-citations are sometimes condemned as a potential means for artificially inflating citation rates and thus strengthening the author's own position in the scientific community (Glänzel et al. 2006). Depending on the purposes of the bibliometric studies, they are frequently excluded in the calculation of specific indicators (van Leeuwen et al. 2003). In the case of brand new indicators such as the h-index or the g-index, a huge debate has been raised on whether or not self-citations should be excluded (Hirsch 2005; Schreiber 2007, 2008).

According to Lawani (1982), self-citations can be classified in two main "genera". On the one hand, the so-called synchronous self-citations, which are the self-citations an author *gives*; and on the other, diachronous self-citations, which are those an author *receives*. This study will focus on the latter, since our major concern is the influence of self-citations on the total citations received and their potential distorting effect on the value of citations as a measure of impact.

There are numerous studies on self-citations in available literature, which deal with different aspects of the topic and focus on different units of analysis. It is possible to study the citations received by a given country from the country itself (country self-citations), those received by a given institution from its own scientists (institution self-citation) (Eto 2003; Iribarren-Maestro 2006; Hellsten et al. 2007) or we can study how often a journal is cited by its own publications (journal self-citations) (Leydesdorff 2008), being an important issue as it can be used to manipulate the impact factor of journals (Krauss 2007; Frandsen 2007).

Self-citations at the document level are usually defined as citations in which the citing and the cited documents have at least one author in common. In other words, a self-citation occurs whenever the set of co-authors of the citing paper and that of the cited one are not disjoint, i.e., when such sets share at least one author (Snyder and Bonzi 1998; Aksnes 2003; Glänzel et al. 2004).

This paper focuses on the study of self-citations at the individual level. For its calculation, two different approaches may be used (Glänzel et al. 2006):

1. Author Self-citations: a direct self-citation for a researcher A occurs whenever A is also co-author of a paper citing a publication by A; or, in other words, those self-citations that one researcher receives from him/herself.
2. Total Self-citations or self-citations at the *document level*: all those citations that a document receives from its authors. It is important to note that under this approach, self-citations embrace a wider concept which includes both author self-citations and co-author self-citations.

Is it necessary to suppress self-citations in bibliometric studies? To address this question the level of aggregation of the units of analysis is an important issue that should be borne in mind. At country level, Glänzel and Thijs (2004a) suggest that there is no especial need for excluding self-citations as they do not represent a major problem. At the meso level, Aksnes (2003) argues in favour of suppressing self-citations; he recommends that the

potential effects of self-citations should be carefully considered before using citations as indicators of scientific impact.

At the individual level, the calculation of bibliometric indicators is a very complex and controversial endeavour (Costas and Bordons 2005). The role of self-citations at this level and their potential effects on citation-based indicators are crucial topics that need to be analysed.

## Objectives

The main objective of this study is to contribute to the understanding of the role of self-citations in the communication process with especial emphasis on the micro-level. For that purpose, a comparison between different approaches in the calculation of self-citations at the individual level (author self-citations, co-author self-citations and total self-citations) has been carried out.

The role of author and co-author self-citations and their relation with different research performance indicators are analysed to shed some light on the behaviour of self-citations on the communication process. Different questions are addressed, such as: is it necessary to suppress self-citations at the micro level? What are the differences between suppressing total self-citations or author self-citations? Could this decision affect scientists differently? Results in two different scientific fields are presented and discussed.

## Methodology and data

This study focuses on the research activity of 715 permanent scientists working at the Spanish National Research Council (CSIC) in the fields of Biology and Biomedicine (388 scientists) and Material Science (327 scientists).<sup>1</sup> The scientific output of scientists during the 1994–2004 period has been obtained from the Web of Science. Documents published by scientists in research stays in foreign centres were also tracked and included in the study.

All articles, notes, reviews and letters have been considered in the analysis and all documents were assigned to scientific fields considering the scientific area of the researchers. Total counting of documents was used and documents published in collaboration by scientists from two different fields were assigned equally to both research fields.

### Main characteristics of researchers

Personal data of scientists were obtained from a personnel list provided by CSIC's Department of Human Resources including specific data for each scientist: year of birth, number of years employed at the CSIC, scientific category and scientific field (Biology and Biomedicine/Material Sciences).

Permanent scientists at CSIC belong to one of the following three scientific professional categories, organised in a hierarchical structure: Tenured Scientists (352 scientists, 49%), which is the basic scientific rank; Research Scientists (185, 26%); and Research Professors (178, 25%), which is the highest scientific category for a scientist at the institution.

Researchers were also classified considering their scientific performance, following the methodology suggested by Costas and Bordons (2007) and Costas (2008). This

<sup>1</sup> The Spanish National Research Council (CSIC) is a multidisciplinary institution organised in eight scientific areas, including both Biology and Biomedicine and Material Sciences.

classification of researchers is based on a balanced combination of three scientific dimensions, including a production dimension, a second dimension based on observed impact and a third dimension regarding international visibility. Following this approach, researchers were classified in Top Class, Medium Class and Low Class, where the Top Class includes those researchers with a high performance rate in at least two of the three dimensions, Medium Class researchers present a medium performance rate in two of the three dimensions, and Low Class researchers show low performance rates in at least two of the three dimensions under review.

### Bibliometric indicators

For each researcher, standard CWTS indicators (Moed et al. 1995; van Leeuwen et al. 2003) were obtained: P (total number of documents), C + sc (total citations including self-citations, with a variable citation window), CPP (Citations per Publication), Median Impact Factor of the publication journals, %HCP (% of Highly-Cited Papers) and h-index (Hirsch 2005).

Special mention is due to indicators of self-citations, shown below, and collaboration measures, included to verify the relationship between self-citations and collaboration previously described (Aksnes 2003; Glänzel and Thijs 2004b).

#### (a) *Indicators of self-citations*

For every scientist three main measures of self-citations were obtained:

- “Total self-citations”, i.e., the total number of self-citations (document level) that all the documents produced by a researcher have received.
- “Author self-citations”, these are the citations that one author receives from his/her own documents.
- “Co-author self-citations”, these are the citations that one author receives from his/her co-authors in the cited document. This indicator is obtained by subtracting the number of “Author self-citations” from the “Total number of self-citations”:

$$\text{Co-authorself-citations} = \text{Total self-citations} - \text{Author self-citations}$$

Finally, the total number of “External Citations”, i.e., those citations produced by researchers different from the authors of the cited paper, were also calculated.

#### (b) *Indicators of Scientific Collaboration*

Concerning collaboration, different indicators were calculated at document and at individual level:

- Average number of authors and centres per document;
- Scope of collaboration, documents were classified in one of the following categories according to the scope of collaboration:
  - No collaboration, i.e., documents produced by a single centre.
  - National collaboration, when two or more centres from the same country are involved.
  - International collaboration, when two or more centres from different countries are involved.

- Total co-authors, it refers to the total number of different co-authors of a given researcher, considering his/her whole scientific output.

**Results**

First, a description of the two fields analysed is presented as a general framework, and results about the chronological evolution and temporal trends of self-citations are shown. The second part focuses on the analysis of self-citations from the individual perspective.

The total output published by CSIC scientists amounts to 18,937 documents; which have received a total of 254,264 citations and 66,593 self-citations (26%). A total of 15,394 documents (81%) included citations, while 12,682 (67%) had at least one self-citation.

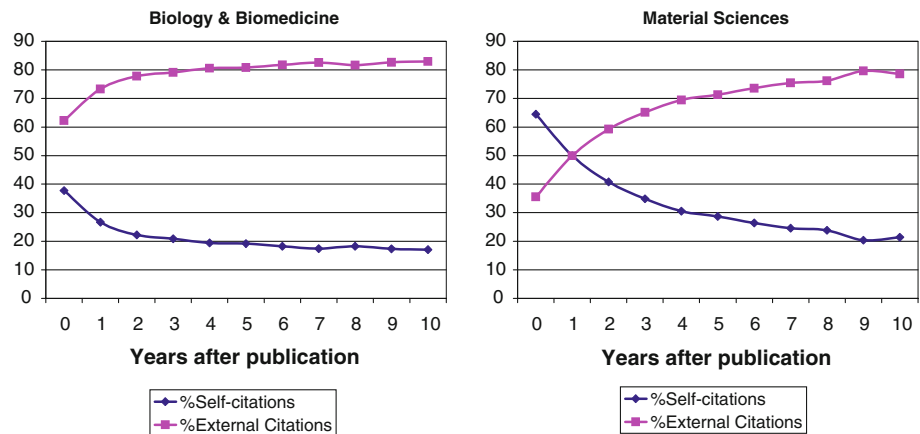
Analysis of self-citations at the document level

The bibliometric description of the two fields under analysis is shown in Table 1. Biology and Biomedicine obtains lower percentages of self-citations and non-cited documents, as well as higher rates of CPP and external citations per document. Inter-field differences were statistically significant ( $p < 0.05$ ).

The evolution of self-citations over time and their relationship with different measures of research performance, such as impact and collaboration are explored in the following paragraphs.

**Table 1** Bibliometric description of scientific fields

Field	P	C	Tot. Self-cit.	%Self-cit.	Cited docs.	%Non-cited docs.	CPP	Self-cit./doc.	External cit.	Ext. cit./doc.	%Ext cit
Biology and biomedicine	9305	177295	38550	21.74	7725	16.98	19.05	4.14	138745	14.91	78.26
Material sciences	9644	77012	28056	36.43	7678	20.39	7.99	2.91	48956	5.08	63.57



**Fig. 1** Temporal evolution of the percentage of self-citations and external citations by scientific field

### Temporal evolution of self-citations

The temporal evolution of self-citations and external citations has been analysed considering the number of years since documents were published (Fig. 1).

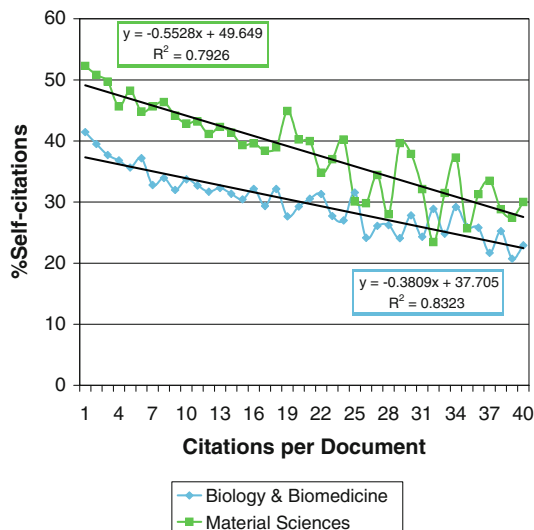
According to Fig. 1, self-citations are more common during the first years following publication, decreasing as time goes by, while external citations are less frequent during the first years following publication and they increase as documents get older. Inter-field differences are evident: documents in Biology and Biomedicine receive from very early stages more external citations than self-citations, while for Material Sciences it is only after the second year following publication that external citations are more frequent than self-citations.

### Relationship of citations and impact factor with self-citations

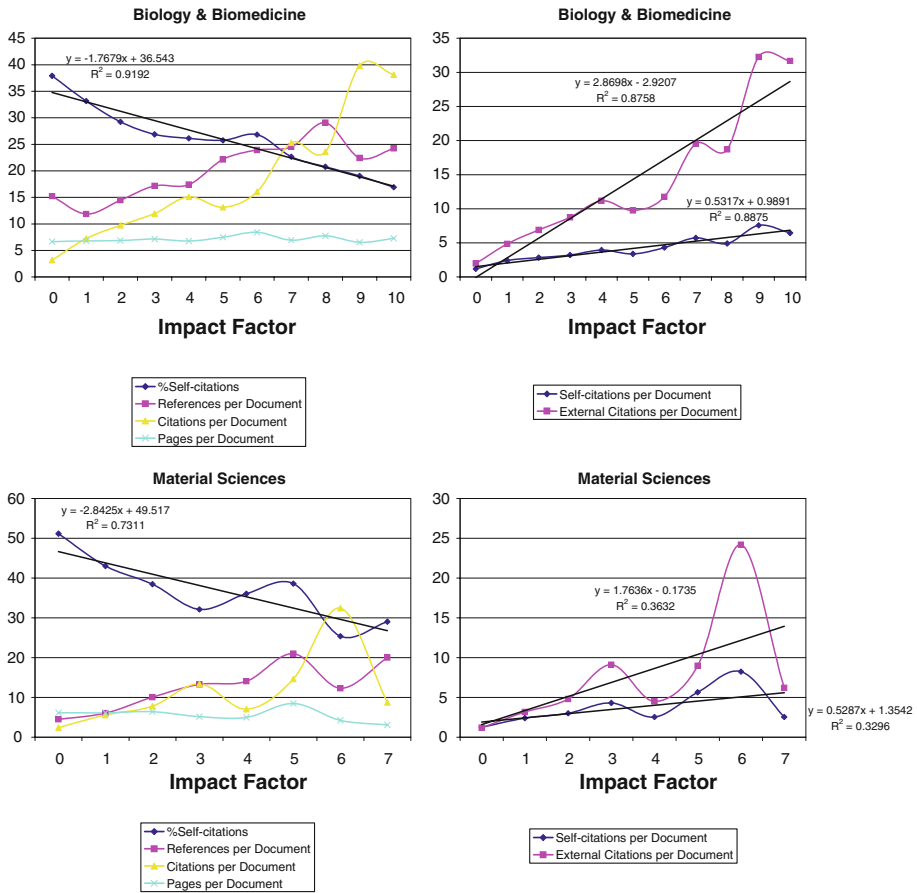
Measuring to what extent self-citations might be decisive in the high number of citations received by some documents is an important aspect that should be dealt with. Our results suggest that self-citations do not play a very decisive role, since documents with more citations tend to present proportionally fewer self-citations, as shown in Fig. 2. This is consistent with the conclusions of previous studies (Aksnes 2003).

Not only the share of self-citations tends to decrease for the highest cited documents, but also for those published in the highest impact factor journals. As a general pattern shared by the two fields analysed we observe that the percentage of self-citations decreases as the impact factor increases (Fig. 3). However, this is not due to a real decline in the raw number of self-citations (actually it grows as shown in Fig. 3-right-) but to a higher rate of growth of external citations.

As the impact factor of journals rises, an upward trend in the number of references and citations per document is observed, while the number of pages remains stable. This means that documents in high impact factor journals are more densely documented (longer references lists) and produce a stronger impact on their community (high number of citations), which is not accomplished at the expense of self-citations.



**Fig. 2** Percentage of self-citations according to the number of citations

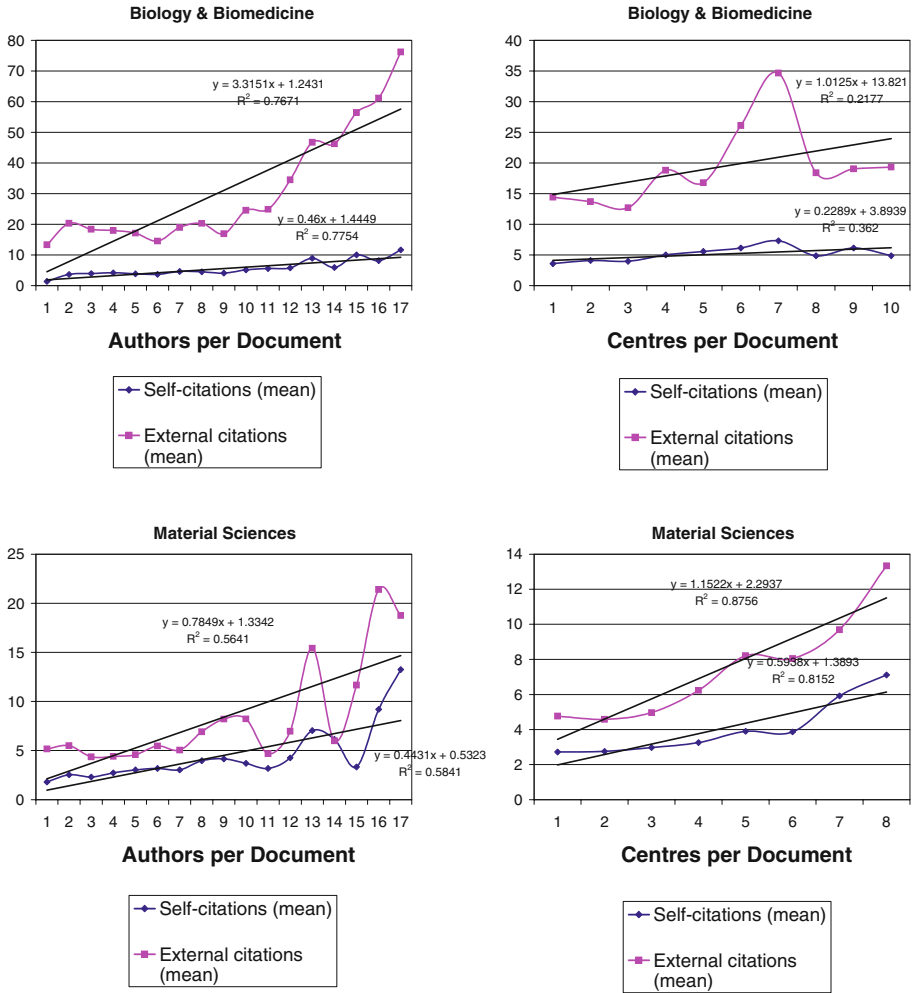


**Fig. 3** Citations, self-citations and external citations by impact factor

*Collaboration analysis*

The role of self-citations in the higher impact described for collaborative documents (see for example Persson et al. 2004) is analysed hereunder. Although some authors suggest that this stronger impact could be due to a higher rate of self-citations (Herbertz 1995) supported by the higher number of authors in documents produced in collaboration, Glänzel and Thijs recently concluded that co-authorship has a strong effect on external citations, but only a moderate effect on self-citations (Glänzel and Thijs 2004b). In Fig. 4, we can see that as the number of authors and centres increases, the number of external citations and self-citations also rises; and the growth of external citations is much higher than that of self-citations. A feature that is especially clear in Biology and Biomedicine, where an additional author results in an increase of 0.5 self-citations and three external citations.

In the case of international collaboration, when compared with the rest of documents, a higher citation rate is observed. As Table 2 shows, documents produced in international collaboration present a higher number of total citations. In fact, both the number of external citations and self-citations are higher for internationally co-authored documents



**Fig. 4** Mean values of self-citations and external citations by number of authors and centres per document

than for those nationally co-authored or with no collaboration at all ( $p < 0.05$ ). Interestingly enough, internationally co-authored documents present a higher number of authors and centres per document ( $p < 0.001$ ) and a higher percentage of self-citations in the two fields under study ( $p < 0.05$ ), a fact that might contribute to their higher citation rates.

**Analysis of self-citations at the individual level**

In this section, the units of analysis are scientists. Only those scientists with a total production greater or equal to the Percentile 10 of their area are included in the study (P10 = 8 documents for Biology and Biomedicine and P10 = 14 documents for Material Sciences). Research performance of scientists at the individual level is described by means of the same indicators calculated above for the description of the field at the meso level, and two new measures concerning self-citations (author self-citations and co-author self-

**Table 2** Self-citations by scientific field and type of collaboration

Field	Type of collaboration	No. citations/document			%Self-citations
		Total citations	External citations	Self-citations	
Biology and biomedicine	No collaboration (3170)	17.88 ± 37.31 7	14.32 ± 34.08 5	3.56 ± 5.34 2	28.80 ± 27.34 22.45
	National collaboration (2977)	16.2 ± 32.99 7	12.68 ± 29.76 4	3.51 ± 5.19 2	30.94 ± 28.28 25
	International collaboration (3168)	22.91 ± 51.8 10	17.59 ± 44.77 6	5.32 ± 9.05 3	34.26 ± 27.31 29.41
	Total (9318)	19.05 ± 41.74 8	14.91 ± 36.94 5	4.14 ± 6.85 2	31.39 ± 27.72 25
Material Sciences	No collaboration (2852)	7.48 ± 17.98 3	4.77 ± 14.53 1	2.71 ± 4.78 1	44.32 ± 35.00 40
	National collaboration (2664)	7.5 ± 13.82 3	4.75 ± 10.71 2	2.75 ± 4.42 1	44.58 ± 34.15 41.18
	International collaboration (4144)	8.65 ± 17.37 4	5.5 ± 14.1 2	3.15 ± 4.91 1	46.49 ± 34.39 44.44
	Total (9660)	7.99 ± 16.67 3	5.08 ± 13.39 2	2.91 ± 4.74 1	45.32 ± 34.51 42.33

Data expressed as: mean ± SD, median

citations) are included (Table 3). The percentages of external citations, author self-citations and co-author self-citations have been calculated for each researcher, considering the total number of citations received. Significant differences by field were found (Fig. 5).

No strong correlation between author and co-author self-citations is observed. Thus, for a given value of author self-citations, different values of co-author self-citations were found depending on the scientist. In other words, these two types of self-citations are not equally distributed among researchers (Fig. 6).

#### *Temporal evolution of self-citations at individual level*

Temporal evolution of external citations and self-citations—considering author and co-author self-citations separately—is shown in Fig. 7.

Figure 7 shows how author and co-author self-citations are more frequent during the first years following publication. However, it is important to note that author self-citations predominate over co-author self-citations, especially during the first 2–3 years following publication; thereafter both author and co-author self-citations tend to converge in both fields.

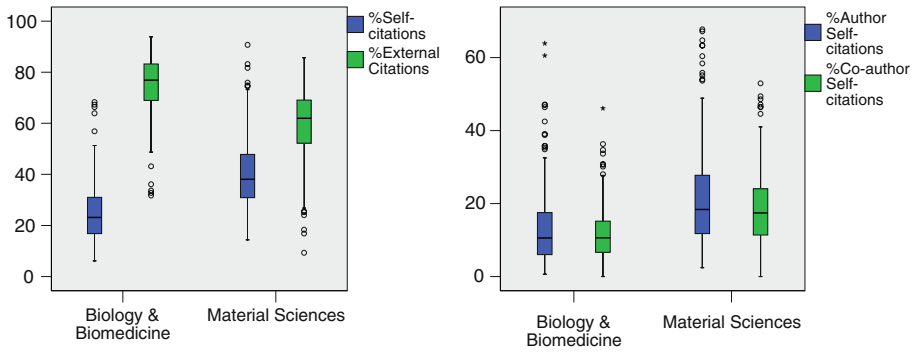
#### *Self-citations by professional category*

The professional category of scientists has been considered for the analysis of self-citations. In Fig. 8a, we can see that there are no differences in the percentage of self-citations and external citations by category. In general terms, from said Figure we may conclude that scientists tend to receive more external citations (around 60–75%) than self-citations (around 20–40%) in all the professional categories and in both scientific fields under study.

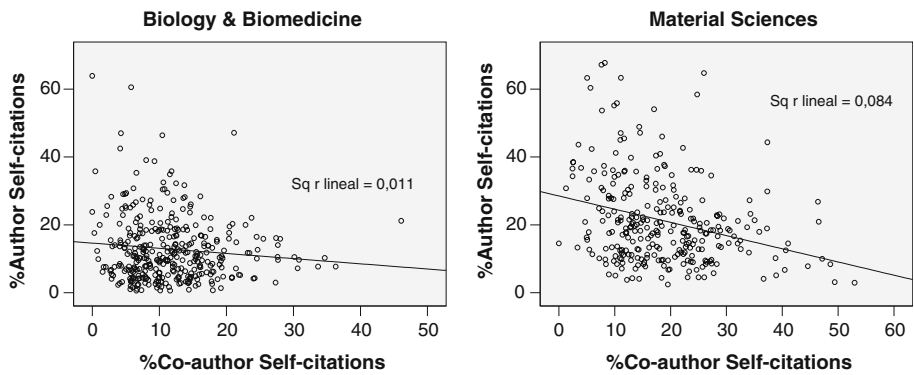
**Table 3** Research performance of individual scientists by scientific field

Field	P	CPP	%Non-cited docs.	%Self-citations	%External citations	%Author Self-citations	%Co-author self-citations
Biology and biomedicine (353)	30.61 ± 20.52 25	21.56 ± 17.58 16.73	14.8 ± 10.57 13.33	24.49 ± 11.03 23.13	75.51 ± 11.03 76.87	12.86 ± 9.54 10.57	11.63 ± 6.65 10.59
Material sciences (284)	53.63 ± 37.22 43.5	7.67 ± 5.38 6.17	34.15 ± 13.45 32.2	39.84 ± 13.68 38.04	60.16 ± 13.68 61.96	21.3 ± 12.91 18.39	18.54 ± 9.61 17.43
Total (637)	40.87 ± 31.32 33	15.37 ± 15.22 10.86	23.43 ± 15.33 21.05	31.33 ± 14.45 29.58	68.67 ± 14.45 70.42	16.62 ± 11.92 13.76	14.71 ± 8.79 12.91

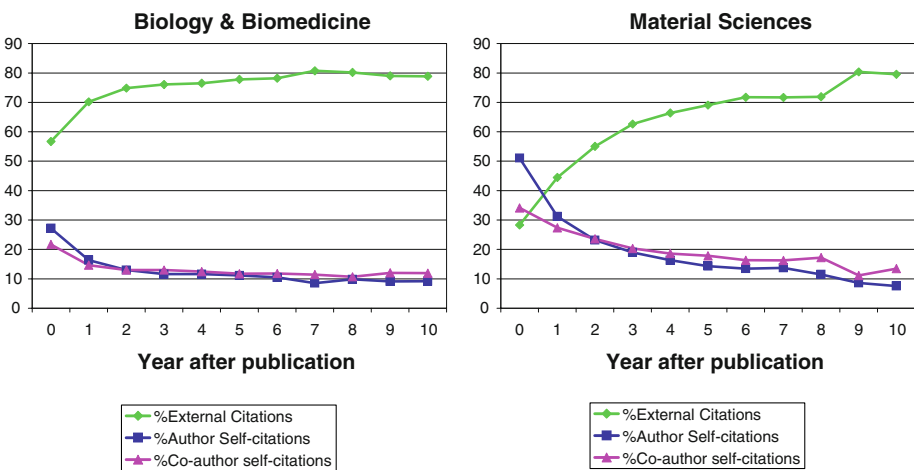
Data expressed as: mean ± SD, median



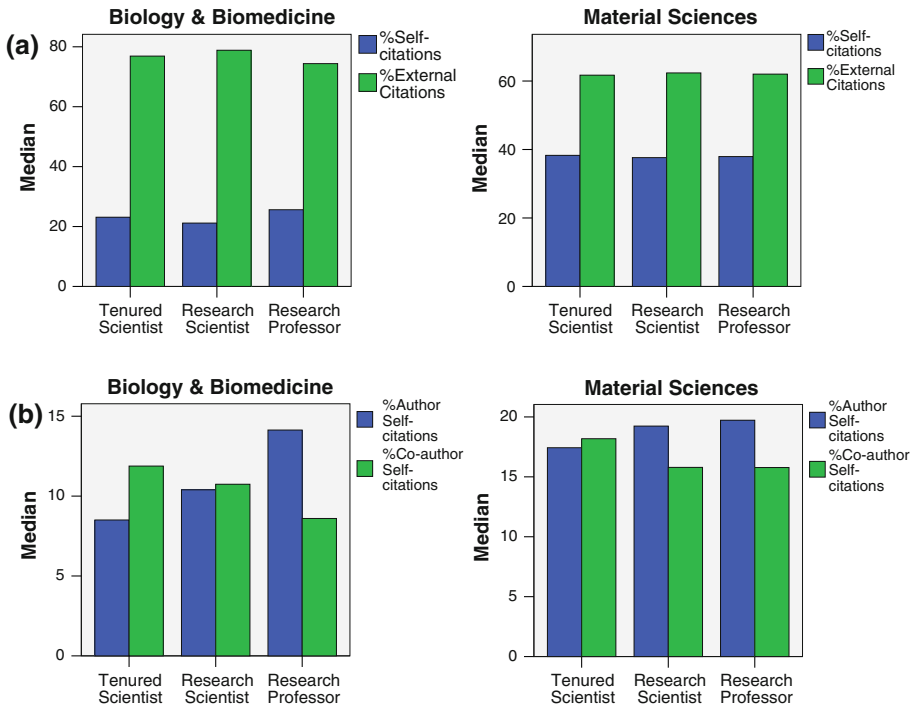
**Fig. 5** Percentage of self-citations and external citations (*left figure*) and author and co-author self-citations (*right figure*) by scientific field



**Fig. 6** Correlations between author and co-author self-citations by scientific field



**Fig. 7** Temporal evolution of the percentage of author and co-author self-citations and external citations by scientific field



**Fig. 8** **a** Percentage of self-citations and external citations by scientific field and professional category. **b** Percentage of author and co-author self-citations by scientific field and professional category

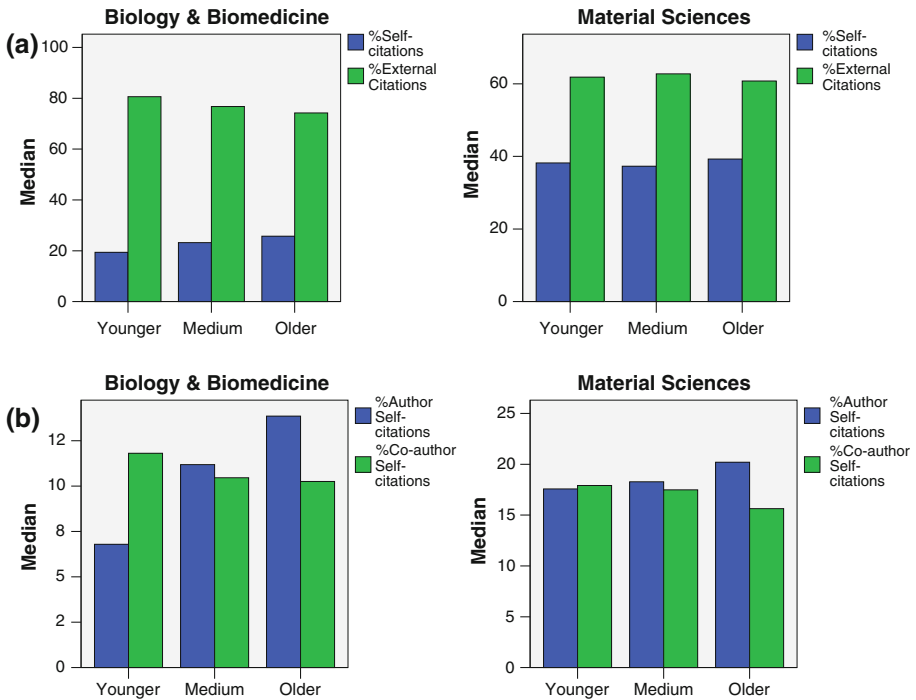
However, when considering the two types of self-citations (author and co-author self-citations) differences among categories come to the fore. In this sense, Research Professors present the highest percentage of author self-citations while Tenured Scientists have the highest rate of co-author self-citations (Fig. 8b). Statistical differences were observed between Tenured scientists and Research Professors ( $p < 0.05$ ).

### *Self-citations by age*

Are there differences in the self-citation share of scientists according to their age? To address this question, scientists have been classified by age in three groups:

- “Younger”: scientists aged 43 or below, that is the percentile 25 of the whole age distribution of scientists;
- “Medium”: scientists aged between 44 and 56 (values between percentiles 25 and 75).
- “Older”: scientists aged 56 or above (percentile 75).

Considering this classification, no differences in the percentages of self-citations and external citations by age were found (Fig. 9a). However, Fig. 9b shows that younger scientists tend to have fewer author self-citations than co-author self-citations, while the opposite holds for medium and older scientists. Statistical differences were found between younger scientists and medium scientists ( $p < .05$ ) in the percentage of co-author self-citations.



**Fig. 9** **a** Percentage of self-citations and external citations by scientific field and age. **b** Percentage of author and co-author self-citations by scientific field and age

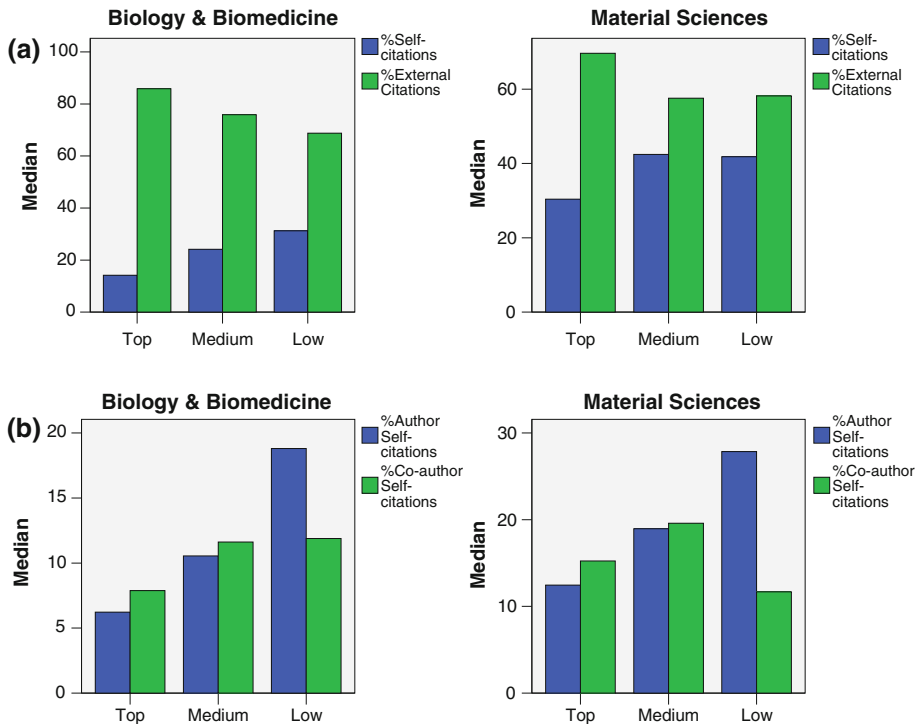
### *Self-citations by scientific class*

The classification of researchers as per their bibliometric performance (Costas 2008) allows us to distinguish three scientific classes: Top, Medium and Low. The percentage of self-citations increases from Top to Low researchers (Fig. 10a). The percentage of author self-citations also increases from Top to Low (Fig. 10b). However, while co-author self-citations predominate over author self-citations for Top and Medium scientists, the latter are more frequent among Low class scientists. This could be related to their lower productivity and collaboration rates.

### *Self-citations, impact and collaboration*

To gain insight on the role of author and co-author self-citations, their relationships with other indicators of the research performance of scientists are studied by means of factor analysis (variables normalised with the natural Logarithm). In both fields, Material Sciences and Biology and Biomedicine, three different components were obtained, which accounted for 80% of the total variance (Tables 4, 5). A global analysis of both areas is presented below, once similar patterns were found in each of the two areas when analysed separately.

The first component reflects relative impact (high-factor loadings for CPP, %HCP and Median Impact Factor), the second is a quantitative-oriented component related to activity and impact in absolute terms (high-factor loadings of total number of documents—P-, total



**Fig. 10** **a** Percentage of self-citations and external citations by scientific field and scientific class. **b** Percentage of author and co-author self-citations by scientific field and scientific class

number of citations— $C + sc$ -,  $h$ -index and the Total number of co-authors); and the third refers to collaboration patterns.

The share of author and co-author self-citations does not substantially contribute to the quantitative-oriented dimension. This notwithstanding, in Biology and Biomedicine the share of author self-citations shows a slight and positive contribution to this dimension, i.e., the share of author self-citations tends to grow with an increasing number of publications (data not shown).

The percentage of total self-citations and author self-citations are highly negatively correlated with the relative impact-oriented dimension in both fields. This is very interesting, since it means that self-citations do not play an important role in obtaining a high relative impact, which is accomplished by means of external citations.

Finally, the third dimension refers to collaboration patterns. The percentage of documents in collaboration and the total number of co-authors contribute to this dimension together with the percentage of co-author self-citations. Thus, the share of co-author self-citations tends to grow for the most collaborative scientists.

#### *Influence of self-citations on the position of scientists in the ranking by CPP*

Do self-citations significantly influence the average CPP values of scientists? As Fig. 11 shows, a strong correlation is found between  $CPP+sc$  (with self-citations),  $CPP-sc$  (all self-citations suppressed) and  $CPP-asc$  (only author self-citations suppressed) in both fields when analysed at the individual level.

**Table 4** Factor analysis in material sciences and biology and biomedicine

Component	Initial Eigen values			Rotation sums of squared loadings		
	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %
1	4.428	40.257	40.257	4.082	37.110	37.110
2	3.162	28.746	69.004	3.160	28.723	65.833
3	1.227	11.153	80.156	1.576	14.323	80.156
4	0.831	7.558	87.714			
5	0.532	4.840	92.554			
6	0.461	4.188	96.742			
7	0.154	1.397	98.139			
8	0.091	0.825	98.964			
9	0.059	0.540	99.504			
10	0.048	0.433	99.937			
11	0.007	0.063	100.000			

Total variance explained

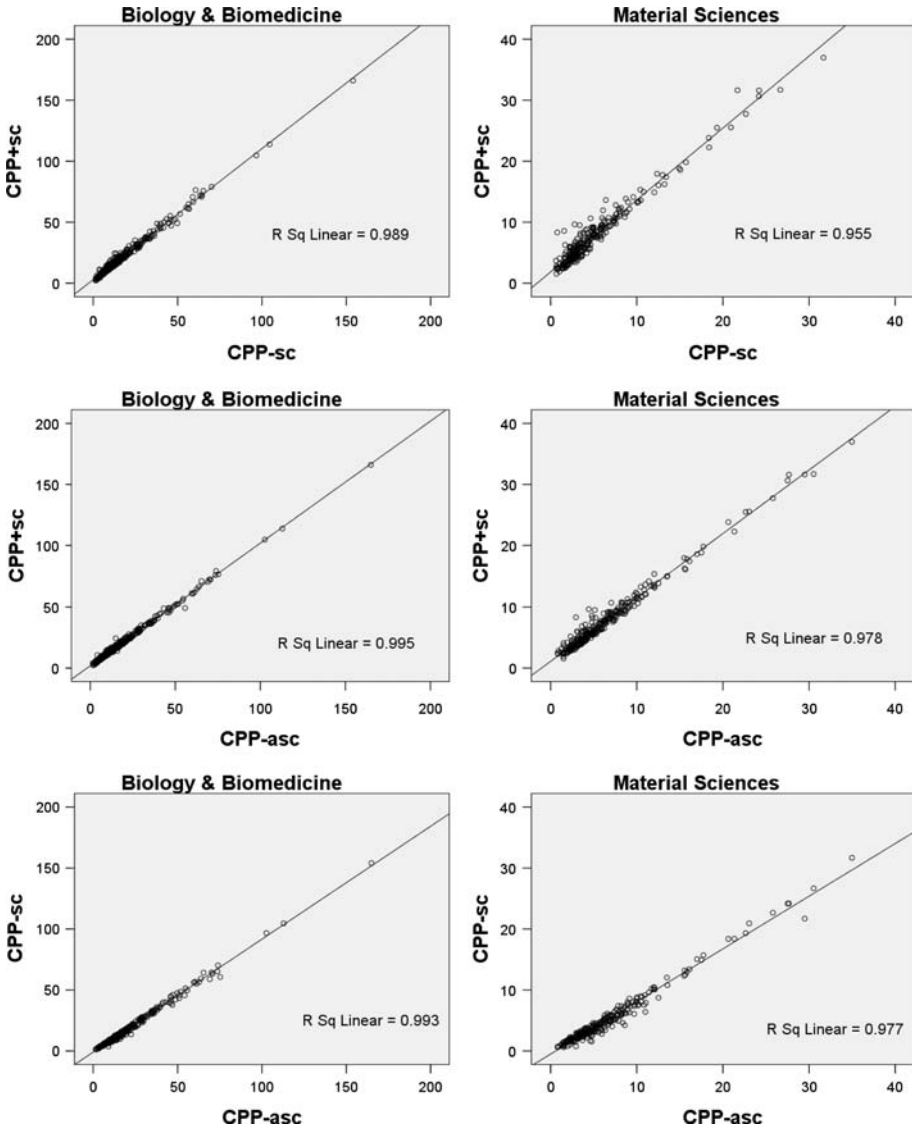
**Table 5** Rotated component matrix in material sciences and biology and biomedicine

	Component		
	1	2	3
CPP	0.933	0.268	-0.119
%Self-citations	-0.880	0.045	0.259
%Author self-citations	-0.867	0.198	-0.206
Median IF	0.817	0.177	-0.067
%HCP	0.622	0.386	-0.065
h-index	0.246	0.933	0.030
C + sc	0.418	0.882	0.019
P	-0.516	0.798	0.148
Total co-authors	-0.036	0.710	0.539
% co-author self-citations	-0.332	-0.093	0.756
% Documents in collaboration	0.125	0.263	0.748

Note: *P* publications, *C + sc* citations (unsuppressed self-citations), *CPP* citation per publication, *%HCP* percentage of highly-cited papers

However, could self-citations influence the position of scientists in the ranking by CPP? To address this question the position of the researchers in the ranking by CPP + sc, by CPP - sc and by CPP - asc were compared. Our findings show that the position of scientists in the CPP rankings may change significantly in one field depending on whether or not self-citations are suppressed (Wilcoxon test in Table 6).

According to the Wilcoxon test (Table 6), in Biology and Biomedicine there are not statistically significant differences in the position of scientists in the ranking by CPP, regardless of the calculation method used (suppressing all self-citations, author self-citations or retaining them all). However, the type of calculation seems to be more decisive in Material Sciences, in which significant differences are found according to the method used. It is interesting to remark that scientists in the first half of the CPP rankings are less



**Fig. 11** Correlations between different CPP scores (individual level). *Note:* CPP citations per publication,  $CPP + sc$  CPP (unsuppressed self-citations),  $CPP - sc$  CPP (all self-citations suppressed),  $CPP - asc$  CPP (author self-citations suppressed)

influenced by the type of calculation than those at the bottom of the list. In Fig. 12 the average difference between researchers' position in two different CPP rankings: a) the  $CPP + sc$  and the  $CPP - asc$  rankings; and b) the  $CPP + sc$  and the  $CPP - sc$ , is shown by quartiles. According to this figure, Material Sciences scientists in Quartiles 3 and 4 of the  $CPP + sc$  ranking are the ones who show the most significant change in their positions when self-citations are suppressed.

**Table 6** Wilcoxon test for the ranks of researchers by CPP + sc and CPP – sc

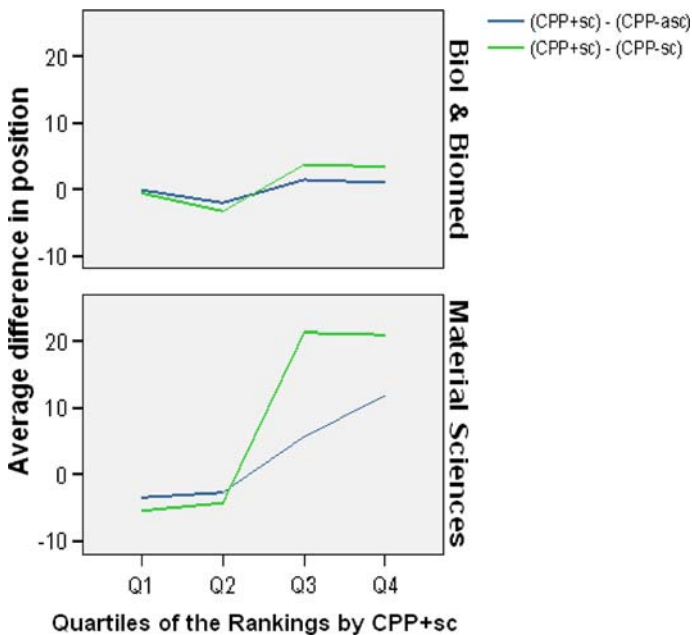
Field		Ranking by CPP + sc vs ranking by CPP – sc	Ranking by CPP – asc vs ranking by CPP + sc
Biology and biomed.	Z	-1.606 <sup>a</sup>	-1.247 <sup>b</sup>
	Asymp. sig. (2-tailed)	0.108	0.212
Material sciences	Z	-3.838 <sup>a</sup>	-2.391 <sup>b</sup>
	Asymp. sig. (2-tailed)	0.000	0.017

<sup>a</sup> Based on negative ranks

<sup>b</sup> Based on positive ranks

Wilcoxon signed Rank test

Rankings for these results are based on integer values of CPP



**Fig. 12** Inter-ranking comparison of the position of scientists (by quartiles)

**Discussion**

Given the cumulative nature of the production of new knowledge, self-citations constitute a natural part of the communication process. Scientists build upon their own results and self-citations represent the use of prior results in the present research. However, in research policy, citations are used as a measure of the impact of research and from this viewpoint self-citations may be considered as a source of distortion.

Different studies conclude that there is no reason for suppressing self-citations at the macro level (Glänzel et al. 2004, 2006; Glänzel and Thijs 2004a) while their potential

effects at the meso level may be more significant (Thijs and Glänzel 2006). On the other hand, their influence at the micro level has been analysed to a lesser degree.

#### Some features of self-citations at the meso level

At the meso level, some of the results obtained in this study are consistent with those described in earlier studies.

- (a) Number of citations and self-citations by field. Inter-field differences in the presence and behaviour of self-citations are explored in this paper. Biology and Biomedicine presents a higher number of citations and self-citations per document than Material Sciences. This is consistent with the higher density of citations (higher *FCS<sub>m</sub>*) described for Biology and Biomedicine when compared with those for Material Sciences at the international level (Aksnes 2003). In fact, Material Sciences outruns Biology and Biomedicine as an applied discipline in terms of the research level of their journals (Morillo et al. 2003) and a higher density of citations has been described for basic fields as compared to applied ones (van Raan 2008).
- (b) Self-citations rate. Biology and Biomedicine presents a lower percentage of self-citations than Material Science. Inter-field differences in the share of self-citations have been described elsewhere and attributed to field variation in citation norms, the extent of cumulative work, and the scope of the field (Aksnes 2003). The fact that scientists in Material Sciences show higher individual productivity than those in Biology and Biomedicine (Costas 2008) might also contribute to this field's higher self-citation rate, since scientists have more recent publications of their own to cite.
- (c) Temporal evolution of self-citations. A faster ageing of self-citations as compared to all citations has been observed in the two fields analysed in this paper. This temporal evolution of self-citations was described for science in the whole world (Schubert et al. 2006) and for different countries (Aksnes 2003) and disciplines (Glänzel et al. 2004). Different underlying reasons for the faster ageing of self-citations can be mentioned. Firstly, scientists themselves are the first ones in using their new findings (self-citations) and only after a period of time has elapsed, their findings are assumed by others in the scientific community (Hellsten et al. 2007). Moreover, different authors suggest that self-citations constitute a means of advertising and disseminating one's own recent work (Medoff 2006; Fowler and Aksnes 2007).  
The inter-field differences found in the temporal evolution of self-citations are one of this paper's interesting results. Thus, within the first year following publication, half of the citations received by Material Science documents are self-citations, while this percentage is below 30% in Biology and Biomedicine. Interestingly, the percentage of self-citations ten years after publication decreases to 20% in both fields. According to Glänzel et al. (2004), self-citations become quite stable 3–4 years after the publication of documents. In our study, this was the case in Biology and Biomedicine, whilst stable values in Material Sciences were obtained much later. Again, differences among fields in the process of production of new knowledge and reporting practices (Hyland 2003), including the ageing rate of literature—which is faster in Biology than in Material Sciences—, may contribute to explain this finding.
- (d) Number of citations and self-citations rises with collaboration. In our study, both the number of external citations and the number of self-citations tend to increase as the number of authors/centres involved grows, but the number of external citations increases at a faster rate. This upholds the results of other authors (Aksnes 2003)

leading to the conclusion that multi-authorship increases above all the probability to be cited by others (Glänzel and Thijs 2004b). Our results show a higher self-citation rate for internationally co-authored documents, a finding that could be related to the higher number of authors and centres involved in these documents, but might not play a relevant role as an “impact amplifier” (van Raan 1998).

- (e) The percentage of self-citations dwindles as the observed impact (citations/document) and the expected impact (impact factor of publication journals) of document increase. This is an interesting finding which suggests that self-citations do not play an important role in the citation rates attained by the highest-cited documents.

### Self-citations at the individual level

The study of self-citations and their relationship with other indicators of research performance at the micro level provides interesting data on the behaviour of the different types of self-citations.

Differences between author and co-author self-citations become apparent. Author self-citations tend to grow, although very slightly, with productivity, probably because very productive scientists have more potential documents to be cited, while co-author self-citations tend to grow very clearly for the most collaborative scientists. A high number of different authors is not always linked to high percentages of co-author self-citations but the link does exist when the percentage of collaboration increases too. Our explanation is that in the latter case different research teams are usually involved (multi-centre documents), and they sometimes collaborate but also work on their own. Therefore, joint-publications can be then cited separately by the different teams involved, resulting in co-author self-citations.

Should we suppress self-citations from citation-based indicators? Our results show that self-citations do not contribute largely to boost the absolute number of citations or the average number of citations per document at the individual level. In fact, high values of relative impact are mainly due to external citations. From this viewpoint, self-citations do not invalidate the use of citations for identifying highly cited scientists.

However, we have observed that scientists in the second half of the CPP ranking may significantly change their positions therein depending on whether or not self-citations are suppressed. Suppressing self-citations is more likely to influence scientists with low CPP values, maybe because they are the ones with the bigger share of self-citations. Therefore, suppressing self-citations could be more adequate for comparing scientists, especially if those with low citation rates are involved.

On the other hand, authors could try to boost their citation rate by self-citing their own documents, but they do not have any control on co-author self-citations. Since the latter are less exposed to manipulation, suppressing only author self-citations—which do not affect all scientists alike—can be an interesting alternative in research assessment exercises.

According to the results presented in this study, external citations and self-citations play specific roles in the scientific communication process. While external citations are the most relevant for evaluative purposes, author and co-author self-citations also provide interesting information concerning the transfer of knowledge.

- *External citations* measure the impact of research beyond the original producers. For evaluative purposes, these citations are the most reliable measure of impact, since they are independent from the producers of the new knowledge, and they can hardly be manipulated.

- *Author self-citations* are very important in the normal process of scientific communication, as scientists need to refer to their previous results as a sign of continuity in their research line. Although scientists may increase their total citation rate by means of author self-citations (manipulative practices), it is not possible to attain high citation rates based only on self-citations. In any event, it is desirable to monitor author self-citations shares and advise against extremely high figures. There are some exceptional situations such as working in new emerging fields or in very narrow or specialized fields in which high rates are justified.
- *Co-Author self-citations* represent the transfer of knowledge among those who produced the original knowledge. It is highly related to the collaborative capacity of researchers, since it tends to grow with the number of collaborators. Collaboration among different teams which also do research on their own may result in an increase of co-author self-citations. For evaluative purposes, they are not as relevant as external citations, but provide meaningful information. Scientists can be advised against author self-citations, but they do not have any control on co-author self-citations.

As a general recommendation for analysts of bibliometric results, the indicator “percentage of self-citations” sometimes presented in reports at the micro-level should always be explained carefully, especially if this measure is prone to be used for detecting anomalous behaviours or endogamy in the self-reference practices of authors. According to our results, indicators of self-citations based on the total self-citations (document level approach) could lead to the notion that individual researchers (or even groups) are responsible for more self-citations than they actually are (especially younger researchers). In this sense, only author self-citations (excluding co-author self-citations) should be considered for evaluation committees if they want to identify unseemly behaviours.

Finally, it is important to note that papers dealing with self-citations at the micro level should mention the methodology used for its calculation: do they refer to total self-citations or to author self-citations? Including both author and co-author self-citations may provide additional information useful not only for research policy purposes, but also to gain insight into the communication process in the research field.

**Acknowledgments** This study was completed thanks to an I3P-CSIC grant at CINDOC (now IEDCYT) and also thanks to a research stay grant at the CWTS in Leiden (The Netherlands). Authors are grateful to two anonymous referees for their comments and suggestions on an earlier version of this paper.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Aksnes, D. W. (2003). A macro study of self-citations. *Scientometrics*, *56*(2), 235–246.
- Costas, R. (2008). Bibliometric analysis of the scientific activity of CSIC researchers in three areas: Biology & biomedicine, material sciences and natural resources. A methodological approach at the micro-level (Web of Science, 1994–2004). Thesis dissertation, Madrid, Carlos III University.
- Costas, R., & Bordons, M. (2005). Bibliometric indicators at the micro-level: Some results in the area of natural resources at the Spanish CSIC. *Research Evaluation*, *14*(2), 110–120.
- Costas, R., & Bordons, M. (2007). A classificatory scheme for the analysis of bibliometric profiles at the micro level. *Proceedings of ISSI 2007 11th international conference of the international society for scientometrics and informetrics* (pp. 226–230). Madrid: CSIC.

- Eto, H. (2003). Interdisciplinary information input and output of nano-technology project. *Scientometrics*, 58(1), 5–33.
- Fowler, J. H., & Aksnes, D. W. (2007). Does self-citation pay? *Scientometrics*, 72(3), 427–437.
- Frandsen, T. F. (2007). Journal self-citations—analysing the JIF mechanism. *Journal of Informetrics*, 1, 47–58.
- Glänzel, W., Debackere, K., Thijs, B., & Schubert, A. (2006). A concise review on the role of author self-citations in information science, bibliometrics and science policy. *Scientometrics*, 67(2), 263–277.
- Glänzel, W., & Thijs, B. (2004a). The influence of author self-citations on bibliometric macro-indicators. *Scientometrics*, 59(3), 281–310.
- Glänzel, W., & Thijs, B. (2004b). Does coauthorship inflate the share of self-citations? *Scientometrics*, 61(3+), 395–404.
- Glänzel, W., Thijs, B., & Schlemmer, B. (2004). A bibliometric approach to the role of author self-citations in scientific communication. *Scientometrics*, 59(1), 63–77.
- Hellsten, I., Lambiotte, R., Scharnhorst, A., & Ausloss, M. (2007). Self-citations, co-authorships and keywords: A new approach to scientists' field mobility? *Scientometrics*, 72(3), 469–486.
- Herbertz, H. (1995). Does it pay to cooperate? A bibliometric case study in molecular biology. *Scientometrics*, 33(1), 117–122.
- Hirsch, J. 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(46), 16569–16572.
- Hyland, K. (2003). Self-citation and self-reference: Credibility and promotion in academic publication. *Journal of the American Society for Information Science and Technology*, 54(3), 251–259.
- Iribarren-Maestro, I. (2006). Producción científica y visibilidad de los investigadores de la Universidad Carlos III de Madrid en las bases de datos del ISI, 1997–2003. Thesis Dissertation, Carlos III University, Madrid.
- Krauss, J. (2007). Journal self-citation rates in ecological sciences. *Scientometrics*, 73(1), 79–89.
- Lawani, S. M. (1982). On the heterogeneity and classification of author self-citations. *Journal of the American Society for Information Science*, 33(5), 281–284.
- Leydesdorff, L. (2008). Caveats for the use of citation indicators in research and journals evaluations. *Journal of the American Society for Information Science and Technology*, 59(2), 279–297.
- Medoff, M. H. (2006). The efficiency of self-citations in economics. *Scientometrics*, 69(1), 69–84.
- Moed, H. F., De Bruin, R. E., & van Leeuwen, T. N. (1995). New bibliometric tools for the assessment of national research performance: Database description, overview of indicators and first applications. *Scientometrics*, 33(3), 381–422.
- Morillo, F., Bordons, M., & Gómez, I. (2003). Interdisciplinary in science: A tentative typology of disciplines and research areas. *Journal of the American Society for Information Science and Technology*, 54(13), 1237–1249.
- Persson, O., Wolfgang, G., & Danell, R. (2004). Inflationary bibliometric values: The role of scientific collaboration and the need for relative indicators in evaluative studies. *Scientometrics*, 60(3), 421–432.
- Pichappan, P., & Sarasvady, S. (2002). The other side of the coin: The intricacies of author self-citations. *Scientometrics*, 54(2), 285–290.
- Schreiber, M. (2007). Self-citation corrections for the Hirsch index. *EPL*, 78(3), 30002.
- Schreiber, M. (2008). The influence of self-citation corrections on Egghe's g-index. *Scientometrics*, 76(1), 187–200.
- Schubert, A., Glänzel, W., & Thijs, B. (2006). The weight of author self-citations. A fractional approach to self-citation counting. *Scientometrics*, 67(3), 503–514.
- Snyder, H., & Bonzi, S. (1998). Patterns of self-citation across disciplines. *Journal of Information Science*, 24, 431–435.
- Thijs, B., & Glänzel, W. (2006). The influence of author self-citations on bibliometric meso-indicators. The case of European universities. *Scientometrics*, 66(1), 71–80.
- Van Leeuwen, T. N., Visser, M. S., Moed, H. F., Nederhof, T. J., & van Raan, A. F. J. (2003). The Holy Grail of science policy: exploring and combining bibliometric tools in search of scientific excellence. *Scientometrics*, 57(2), 257–280.
- Van Raan, A. F. J. (1998). The influence of international collaboration on the impact of research results. *Scientometrics*, 42(3), 423–428.
- Van Raan, A. F. J. (2008). Scaling rules in the science system: Influence of field-specific citation characteristics on the impact of research groups. *Journal of the American Society for Information Science and Technology*, 59(4), 565–576.