Scientometrics (2010) 83:337–354
DOI 10.1007/s11192-009-0088-9

Why with bibliometrics the Humanities does not need to be the weakest link

Indicators for research evaluation based on citations, library holdings, and productivity measures

A. J. M. Linmans

Received: 28 January 2009 / Published online: 13 August 2009

Abstract In this study an attempt is made to establish new bibliometric indicators for the assessment of research in the Humanities. Data from a Dutch Faculty of Humanities was used to provide the investigation a sound empirical basis. For several reasons (particularly related to coverage) the standard citation indicators, developed for the sciences, are unsatisfactory. Target expanded citation analysis and the use of oeuvre (lifetime) citation data, as well as the addition of library holdings and productivity indicators enable a more representative and fair assessment. Given the skew distribution of population data, individual rankings can best be determined based on log transformed data. For group rankings this is less urgent because of the central limit theorem. Lifetime citation data is corrected for professional age by means of exponential regression.

Keywords Bibliometrics · Humanities · Citation analysis · Library holding analysis · Research evaluation

Introduction

Bibliometrics in Humanities and sciences

The growing demand for metrics as an aid in assessing the accomplishments of scientific researchers on different levels (individuals, research groups, departments, universities, countries, etc.) also has an effect on the Humanities. Hitherto this area of scholarship has not been endowed with a recognized, widely discussed, let alone endorsed, set of scientometric and bibliometric indicators. This is in contrast to the position that bibliometrics over the last 20 years has gained in the sciences, where particularly citation analysis has been a successful provider of impact indicators with Thomson Reuters' ISI *Science Citation Index* (*SCI*) as the main source of citation data.

A. J. M. Linmans (✉)
Centre for Science and Technology Studies (CWTS), Leiden University, Leiden, The Netherlands
e-mail: linmans@xs4all.nl

In 1975, ISI launched the *Arts and Humanities Citation Index* (*A&HCI*), but so far this index, and more recent indexes like Elsevier's SCOPUS, have played no major role in the performance evaluation of research in the Humanities (Nederhof 2006). Examining the current situation, one gets the impression that the Humanities has a preference for 'soft' metrics. An example of soft metrics I would call the *ad hoc* weighting of scientific output with the help of a given classification of scientific journals as to their importance, e.g. the ERIH lists (ERIH standing for European Reference Index for the Humanities). For an early attempt in this direction see Luwel et al. 1999.

This could in the future lead to two parallel evaluative systems, one for science, and one for the Humanities—a new embodiment of the 'two cultures' described by Snow (1964). Critical of such a development, our study tries to work out an alternative approach, incorporating 'hard metrics' even for the Humanities as much as possible. However, given the specific nature of the Humanities, the indicators needed here will most of the time not be a mere copy of existing science indicators.

This paper was prepared by a preliminary report (Linmans 2008), which contained fuller details on the bibliometric profiles of the individual disciplines. Elementary descriptive statistics will suffice to achieve our present aims. I hope to publish soon a second, more theoretical article on the underlying mathematical functions and distributions. There it will be shown that especially the exponential growth function and the log-normal distribution of performances over sources may provide a more fundamental understanding of the bibliometric laws at work. These findings will be of consequence also to the bibliometrics of science.

Three-level approach: citations, library holdings, and productivity

Our first indicator measures citation impact with the help of the Web of Science (*WoS*) *A&HCI* indexes, covering about 1,000 journals. Citation impact indicators currently used for science are not adequate when it comes to assessing accomplishments in the Humanities. This is first and foremost because standard analyses use only the bibliographic references *fromWoS* source journal publications referring *to* other *WoS* source journal publications. References to non-source publications, though available, are not used. This works well for the sciences, since their source coverage by *WoS* is high. However, with the Humanities, the non-source citations numerically exceed the source citations. References in *A&HCI* source journals refer mainly to non-source journal publications, or to monographs and edited books, i.e. publications by definition excluded as *WoS*-sources.

Consequently, for the Humanities citation analysis has to expand its target material by including non-source citations, if it will have any chance of being representative, fair, and relevant. By taking this course for my data sets, the citations available were increased by factor five. Before, Butler and Visser (2006) have applied the target expanded method to social sciences and humanities, measuring the citation impact of nine Australian universities; see also Moed 2005, pp. 147–157.

Here I will propose a different target expanded approach, which may circumvent some of the more cumbersome aspects of the Butler–Visser method. The main difference is that I will no longer count *citations per publication*, but *citations per author*. Therefore the focus will be on oeuvres. We might refer here to co-citation analysis, where author studies based on oeuvre have a long tradition in bibliometrics. The *h* index too was originally introduced by Hirsch (2005) as a measure of author performance. I will, moreover, use lifetime citation data (i.e. all citations hitherto received by an author), instead of the restricted time windows of standard analyses. Both of our strategies serve to make more data available at a time.

Even so, citation analyses retain severe limitations. Particularly, *WoS* exclusion of citing references in books is a limiting factor. This is a possible source of distortion to reckon with, because there is some evidence that the citation patterns of books might differ from those of journals (Line 1979; Cronin and Snyder 1997). A further deficiency is the Anglo-American bias of the current citation databases (Archambault et al. 2006).

Trying to mend these shortcomings, I will introduce a second group of impact indicators, measuring the extent to which books of the same authors are represented in the collections of representative scientific libraries in different countries. Torres-Salinas and Moed (2009), studying library holdings of books deriving from Spanish universities, are among the first to have used this new type of library holding analysis. Here I will use the catalogue data of libraries in the USA, the UK, and the Netherlands, assembled in the *WorldCat*. Library collection analysis has not only the advantage that it specifically elucidates the impact of books; it will, to a certain extent, also make it possible to correct the Anglo-American bias of citation analysis.

Readers might object that library holdings do not reflect the immediate impact of publications on researchers, but that they are a matter of library policies. To some extent this is true. But on the other side library holding analysis has the great advantage that libraries buy new books during the first years after publication, whereas citations accumulate over a much longer period, in particular with the Humanities.

Finally, to enable comparison of impact with productivity, I will add a simple productivity indicator, based on the annual number of pages published per year. It is unsophisticated, but it is only meant to provide some background reference for the citation and library holding impact indicators, which remain, by all means, the core of the investigation.

Limitations

Admittedly, the indicators put forward cannot be expected to supply under all circumstances a representative picture of each researcher of their research accomplishments. Particularly the lack of specific impact indicators for activities on the Internet may be seen as a serious deficiency, and is certainly something to be mended in the future. Google Scholar already supplies citation data for books which are bound to become a rich mine of bibliometric information in addition to the existing citation databases. However, in order to make Google Scholar a major reliable bibliometric source, one will have to acquire more knowledge on the input policies and practices of Google Scholar, which so far lack transparency.

Still other bibliometric and scientometric indicators might be proposed to help supplement our indicators. Libraries circulation data has been analysed since the 1980s for library housekeeping purposes (e.g. Burrell 1990), but there is no reason not to use this data for impact evaluation as well. Book sales are another candidate. And of course there is a lot of scientometric, non-bibliometric data of possible interest, both on the input side (funding, research hours, etc.), and on the output side (awards and other esteem indicators, economic benefits, social impact, media presence, etc.).

The Leiden data sets

The empirical substratum of this study is formed by a group of 292 faculty members of the Faculty of Humanities at Leiden University (The Netherlands), active in the fields of linguistics, literature, history, art history, supplemented with researchers in the field of anthropology and area studies (partly belonging to the Faculty of Social Sciences). Most of them were university employees by June 2006; a small part of them retired or accepted a post elsewhere between 2002 and 2006. In the citation analysis the data of 264 researchers was used (since for 28 out of the 292 researchers no citations were found in the citation index); for library holding analysis and productivity analysis smaller sub-samples were used for 80 and 62 individuals, respectively, in order to limit work-load (and in the case of productivity because of the incompleteness of data).

The data was derived from *A&HCI* and the *Social Sciences Citation Index* (*SSCI*) for the citations, from *WorldCat* for the library holdings, and from *METIS*, a Dutch database covering research output, for the productivity data. Counts were finalized in January 2007. Citations were counted as citations per author. Library holdings were counted as the number of libraries owning a copy of a specific book. The productivity scores were computed as numbers of pages published per year by each author (annual average over the period 1995–2006).

The key figures for all the data sets are provided in a separate box *Key figures regarding Leiden data sets.*

The data were divided into five disciplines: Linguistics, Literature, History, History of Arts, and Anthropology with area studies. Subject terms assigned by the ISI citation indexes could not be used, because cited non-source publications, the greatest part of our citation data, are not subject-indexed in the *A&HCI.*

Key figures regarding Leiden data sets (January 2007)		
Citations (*Sources*: *WoS*, *A&HCI*, *SSCI*)		
Cited authors	264	
Mean professional age	23.4	
WoS source publications by authors	837	
Source publications per author	3.2	
Source publications per author per year	0.1	
Citations (to *WoS* source publications)	1,468	
Citations (to source and non-source publications)	7,216	
Citations per author (lifetime)	27.3	
Citations per author per year	1.2	
Library holdings (*Source*: *WorldCat*)		
Authors	80	
Books	1,135	
Books per author	14.2	
(Co-)author		59%
(Co-)editor		41%
Books per language		
Books in English		45%
Books in Dutch		45%
Books in other languages		10%
Library holdings	59,386	
Library holdings per book	52.3	
Library holdings per author	724.3	
Library holdings per country		
Library holdings USA		76.3%
Library holdings UK		4.1%
Library holdings Netherlands		19.6%

Productivity (*Source*: *METIS*)		
Authors	62	
Pages (co-)authored and published per year	115.3	
Published in monographs		61.8%
Published in book chapters		24.6%
Published in journals		13.6%

Assessment indicators for the humanities

Citation analysis

Shortcomings of standard citation analysis for the Humanities

Factors, most of them already hinted at, hindering the application of standard citation analysis in the Humanities, are: (i) poor coverage by *A&HCI* of the total literature output of the Humanities; (ii) the discarding in standard citation analysis of bibliographic references to non-source publications; (iii) the use of relatively small time windows, reducing the volume of citations available; (iv) the scarcity of multiple-author publications in the Humanities, which prevents them from benefiting from its citation-multiplying effect; (v) the use of the research group in standard analysis as the aggregate level *par excellence*, while research in the Humanities is overwhelmingly individualistic. I will now make a few additional remarks concerning some of these deficiencies.

Coverage The low coverage by the *WoS* citation indexes of Humanities research literature has been pointed out many times before. Moed (2005, pp. 119–136), using a method of 'internal coverage' measurement (based on the within-index proportion of source and non-source citations), found for many science disciplines coverage figures between 60% and 90%, when compared to 15–35% for the Humanities. Lindholm-Romantschuk and Warner (1996) found for philosophy a ratio 8:1, against 2.5:1 for economics and sociology (based on *WoS* data). Contrary to widely held beliefs, in a recent study Larivière et al. (2006) reported evidence that in the nineties the share of books increased, rather than declined. Examining the publications lists of 62 Leiden researchers, I found that only ca. 10% of their journal publications were source-indexed by *WoS*. When books were also included, I found that only 1.5% of the total output is source-indexed (output measured in pages).

Coverage levels are related to the book: journal ratio, but also to language, geography, audience, and text type. Current data bases have a low coverage of publications in languages other than English (Van Leeuwen et al. 2001), publications with a regional focus and radius, and publications addressing a lay public (instead of a closed community of scientific specialists). Large parts of the activities in the Humanities are thereby lost out of sight.

Lifespan of literature De Solla Price (1970) has pointed out the difference between 'hard science', typically showing a dominant research front, and 'soft science', where archival aims blend in. The pace of science is faster and more compact than that of the Humanities, where knowledge develops slower, in spiralling and often erratic patterns. This gives the literature of the Humanities a longer life cycle. This leads to more protracted citations histories, whereas science disciplines have their citation peak typically within 3 or 4 years after publication. Different types of measurements have been used to quantify the difference, e.g. Price's index (the percentages of references in publications referring to literature not older than 5 years), cited half-life (the age median of citations received), and exponential citation decay.

Evidently, the lifespan of literature in different disciplines is associated with their book : journal ratio, since books are the slower medium.

The multiplier-effect of multiple authorship Not only does data base coverage determine the citation level of a discipline, but also the extent of multiple authorship, since the number of citations available increases as the average number of authors per publication increases. Figure 1 (data courtesy of Henk Moed) illustrates the huge difference in co-authorship numbers that exists between the Humanities and the sciences. Journal publications in the Humanities have an average of 1.06 authors per article and this figure is constant over the years. For the sciences, the figure is much higher, and has been increasing over the last two decades.

Ranking of individuals

Now we will have to see how for the Humanities expanded citation analysis, using both source and non-source citations for whole oeuvres, can solve some of the problems posed by standard citation analysis. First we will have a look at the ranking of individuals. Later, in Section "Ranking of groups", I will pay attention to the specific problems presented by the ranking of groups.

Professional age correction When ranking individuals according to their lifetime citation performance, we must find a solution for the problem that the lifetime data used belongs to researchers of different ages. Without correction, older researchers would be at an advantage over their younger colleagues, since they have accumulated citations over a longer time period.

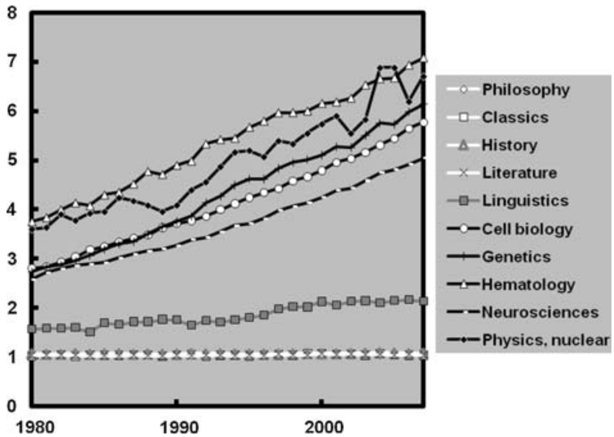

Fig. 1 Number of authors per paper in various disciplines (1980–2007). The Humanities disciplines barely exceed one author per paper. In contrast to the other disciplines they show no clear co-authorship increase over the years. 'Linguistics' refers to journals indexed in the Social Sciences Citation Index and represents the behaviour oriented type of linguistics (data courtesy of Henk Moed)

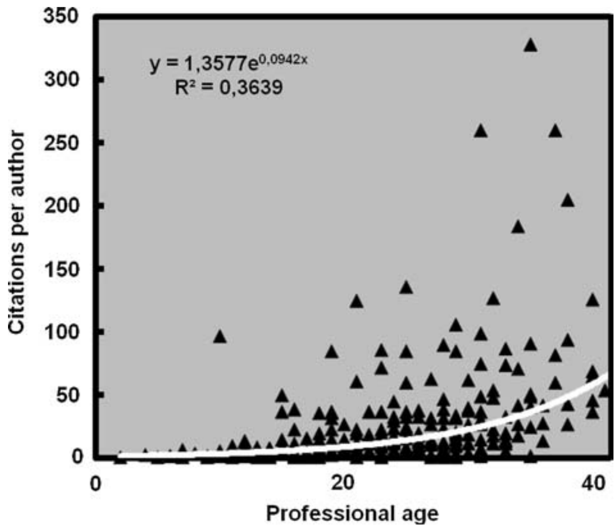

Fig. 2 Scatter plot of citation rates over *PA* of 264 Leiden authors in the Humanities. Exponential regression, represented by the white line, fits the data better than linear regression. Unfortunately, the accumulation of data-points at the bottom visually blurs the density pattern

This is corrected by taking into account the *professional age* (per 2006) of each researcher. In this paper a researcher's professional age (abbreviated *PA*) is defined in terms of the number of years passed since the earliest year the researcher is mentioned in the citation index (which is normally the publication year of the oldest cited publication). On average, researchers were found to have their 'professional birth' around the biological age of 30.

Correction takes its cue from Fig. 2. This figure represents the citation scores of 264 Leiden researchers as a function of *PA*. The figure also represents the exponential regression.

Two types of correction are considered: correction by linear regression and correction by exponential regression. The difference between the two corrections is best demonstrated, when the exponential correction is transformed into a linear correction by taking logs of the citation rates. Then we have two linear correction models: (i) the linearly corrected original (non-log) citation rates; (ii) the linearly corrected log citation rates, which are ultimately, i.e. when seen from the perspective of the original counts, exponentially corrected rates.

In this paper I have always used natural logarithms (to base e). The expression log should therefore always be understood as ln.

On this basis we can establish for each author two standardized, *PA*-corrected scores: the linear score z_k, based on (i); and the exponentially based (but at face value linear), score log $z_{k,}$ based on (ii). The index k indicates the *PA* of the author and it is on this account that the correction is calculated. For mathematical details the reader is referred to the Appendix.

The log space, in which the second score takes its place, is represented by Fig. 3, the logarithmic mirror of Fig. 2.

Exponential regression of the non-log data appears to provide a better explanation of the variation than the linear regression of the non-log data, as can be seen by their coefficient of determination, $R^2 = 0.364$ and $R^2 = 0.204$, respectively. This tells us that citations, as

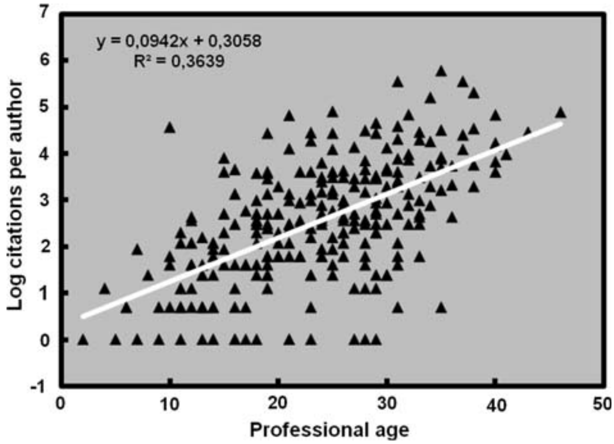

Fig. 3 Scatter plot of logarithmically transformed citation rates for 264 authors in the Humanities. The exponential regression of Fig. 2 is translated into a linear regression

PA increases, do not grow linearly, but that their growth rate is accumulating over the years (like compound interest). Therefore it can be said that the score log z_k, which accounts for exponential growth, is superior to the score z_k, which does not.

In the next section it will be confirmed that the two models sketched here behave in different ways. This is seen at best in the varying relationships in each case between the central tendencies: arithmetic mean, geometric mean (GM) (that is $\left(\prod_{i=1}^{n} x_i\right)^{1/n}$, or the *n*th root of the product of *n* numbers), median, and mode. It can also be seen in the symmetric shape we get when using log z_k, when compared to the skew shape we have when z_k is used. We have to see why this is and what the implications are, not only theoretically, but also for the practice of performance evaluation.

Central tendencies in skew distributions A characteristic feature of lifetime citation rates is that their distribution over researchers is very skew: many researchers with low citation rates contrast with few researchers. This has consequences for the central tendencies.

In skew distributions the median is often much closer to the GM than it is to the arithmetic mean (AM; in everyday language: *the* mean or *the* average). This is confirmed by the Leiden citation rates (Y), where AM(Y) = 27.3 citations per author (cpa), GM(Y) = 12.3 cpa, and Median(Y) = 13 cpa. Mathematically we know that always GM(Y) = AM(log Y) and Median(Y) = Median(log Y). Therefore, since Median(Y) $\sim$ GM(Y) is an empirical characteristic of the Leiden data, it follows immediately that, for the log data, Median(log Y) $\sim$ AM(log Y). This means that, if we replace the original citation frequencies by log frequencies, the AM, having been at a far distance from the Median in the original space, virtually coincides with it in the log space. In log space, the number of researchers above average is now about the same as the number of those below average. In other words, by taking logs, we have translated the original skew data into data divided into two virtually equal halves.

The equal partitioning of log Y by AM(log Y) is not the only symptom of an underlying symmetrical structure. This is best seen, if we take an overview of the spread of the log z_k-scores. These scores range, quite symmetrical, from −2.9 to 3.2. Equally, the lower and upper quartiles are −0.6 and 0.6. The percentages of researchers above and below AM are about equal: 52% and 48%, respectively. If, on the contrary, we take non-log z_k-scores, a very asymmetric picture emerges. The scores range from −1.38 to 7.42; the lower quartile is −0.5, compared with the upper quartile which is 0.15. Above AM we find only 36% of the researchers, below AM 64%.

Which of the two options should be preferred for practical assessment purposes? The aesthetic appeal of symmetry alone cannot be decisive here, nor the fact that statistics have assembled vast knowledge on symmetric (normal) distributions. Some people might even argue that symmetry-inducing log scores are an artefact, blinding us from skewness in the real world, for which it is characteristic that there are few individuals at the top, with great gaps between them, whereas at the bottom it is crowded and the citation frequencies are tight (compare Leydesdorff and Bensman 2006).

There are however three concrete reasons why, when ranking individuals, we should prefer the log z_k-scores of the exponential model over the z_k-scores of the linear model:

1. The exponential growth model fits better, as seen above;
2. The linear model can be misleading, in so far as observers may naively be induced to believe that AM separates researchers in two equal halves;
3. The linear model excessively promotes younger researchers: for the above median scores of the linear model the average *PA* is 20.6 years, whereas the average below median is 26.3 years. While with the exponential model the spread is much more balanced: then the average *PA* above median is 23.9 years, whereas the average below median is 22.9 years.

If one should nevertheless wish to insist on using non-log scores, then one is advised to use the GM as the pivotal central tendency, instead of the arithmetic mean. For the mathematical details, the reader should refer to the Appendix.

Ranking of groups

The ranking of groups is a different story. We will use the word ‘group’ for sets to be ranked on the basis of some data category associated with the set members: research groups in the first place, but also higher level structures (faculties or departments), disciplines, gender, countries, etc. Even individual authors might be approached as ‘sets’ of publications.

Group ranking mainly uses one-figure statistics, particularly the average (i.e. the arithmetic mean). Distributions convey more information, but they are more complex, because they normally have at least two parameters. The average is a good candidate because of the central limit theorem (CLT). This fundamental law of statistics tells us that if we take samples from a population with fixed, though not necessarily known, distribution, the sample means tend to have a normal distribution, no matter what is the specific nature of the population distribution. The approximation becomes better, the larger the sample size is.

There is however a *caveat.* CLT requires sufficiently large samples to make the approximation satisfying (see e.g. Dekking et al. 2005, pp. 195–205). Especially skewed population distributions have the problem of slow convergence. Now this combination of skew population distribution and relatively small sample sizes happens to be our problem. With research groups as our target, we have to work with a fairly limited number of publications. And sample sizes are further scaled down by the use of restricted time windows. Moreover, sample size is far from constant. All these factors prevent an ideal working of CLT.

In order to illustrate how the skewness of the population distribution can linger on in the sampling distribution of group means, I make use of Van Raan (2006). I found that the average citations per publication of 157 chemistry groups presented there (Van Raan Fig. 4) prove to have an approximately lognormal distribution ($R^2 = 0.995$). The same study lets us also observe that the skewness of group means is not removed by field normalization. If we take, for instance, Van Raan's Fig. 13 (presenting field normalized group means *CPP/FCSm*), its lognormal distribution is still unmistakable ($R^2 = 0.991$).

Does it really matter, that the group means distribution is skew? Possibly not as long as one is aware of the asymmetry of the group means, and as long as one realises that this drives a wedge between mean and median. But in situations (often occurring in daily life) where 'above average' and 'below average' are naturally understood as 'belonging to the upper half' and 'belonging to the lower half', misunderstandings will easily arise. The problem can be avoided, if, in the same way as with individual rankings was done in Section "Central tendencies in skew distributions", we decide to use log data. The group mean is then the mean of the log z_k-scores of the individuals of the group, to be denoted as $\overline{\log z_x}$. The CLT approximation is now much smoother, since the underlying data are no longer skew.

Taking into consideration that we can use the data with or without *PA*-correction, we may thus distinguish four types of group mean: (i) $\bar{z}$,the mean of standardized uncorrected *Y*; (ii) $\bar{z}_x$, the mean of standardized, linearly corrected *Y*; (iii) $\overline{\log z}$, the mean of standardized uncorrected log *Y*; and (iv) $\overline{\log z_x}$, the mean of standardized, linearly corrected log *Y*. For more detailed mathematical information see the Appendix.

Table 1 presents the outcomes for the Leiden data set. For each discipline the four group means were calculated. The overall variance reducing effect of taking group means causes the score patterns of the four averaging options to be quite similar. For practical purposes any of the four options may be suitable to give us a rough ranking measure. This is in contrast with individual performance ranking, where it was seen that exponential *PA*-correction was essential.

Although the table shows that the standardized group scores for all score types are roughly similar, switches in ranking position are not excluded (as is shown by History). Unfortunately, there are not enough disciplines in our table to make it directly visible that the non-log *Y* group means are more prone to skewness than the log *Y* means (as with the Van Raan data mentioned above). Nevertheless, the theoretical advantage of using log means should be upheld on the aforementioned grounds.

One aspect that might sometimes contribute to preferring one method over the others is the effect of variance. Taking logs reduces variance. Consequently, the use of log data puts a premium on group homogeneity (low variance). By contrast, groups with great internal

Table 1 Discipline means for citations per author (*Y*): with *PA*-correction (b, d) and without (a, c)

	(a) $\bar{Z}$	(b) $\bar{Z}_x$	(c) $\overline{\log z}$	(d) $\overline{\log z_x}$
Linguistics	0.20	0.19	0.14	0.12
Literature	−0.28	−0.25	−0.30	−0.28
History	0.21	0.18	0.31	0.31
Art history	−0.34	−0.34	−0.48	−0.54
Anthropology	0.27	0.28	0.35	0.41

Author scores, on which the discipline means are based, are standardized (mean 0, variance 1). Scores (a) and (c) have no *PA*-correction; (b) corrects using linear regression; (d) is ultimately based on exponential regression, which is however, by taking log frequencies, transposed to linear regression variety (high variance), caused for instance by exceptionally successful group members, benefit from using non-log data.

Library holding analysis and productivity analysis

The next step will be to add the library holding indicators and productivity indicators.

In library holding analysis book *titles* are the 'types' and book *holdings* (library copies) are the 'tokens'. The average number of library holdings per title per researcher (or group of researchers) determines the rank position of each researcher (or group).

For a subset of 80 authors I found in *WorldCat* (using data of USA, UK, and Dutch libraries) 1,135 scholarly book titles, corresponding with 59,386 book holdings. This gives an average of 52.3 holdings per title (for further summarizing figures see the Key Figure Box). New editions and translations were subsumed under the title of the original publication.

In order to get a more detailed view, also partial impact indicators were constructed by creating the following subsets: (i) library holdings for books in English, with authorship involvement; (ii) *idem*, with editorship involvement; (iii) library holdings of books in Dutch.

To complete the picture we add the productivity indicator. This is calculated as the annual production by number of pages published. In the dataset the results of a subset of 62 researchers were calculated, counted over a period of 10 years. On average these authors published annually 115.3 pages (only authored or co-authored pages were included in calculating this number; edited publications were excluded).

The frequency patterns emerging in library collection analysis and productivity analysis prove to be very similar to those encountered earlier in citation analysis. Table 2 shows that the new data are no less skewed than the citation data, and that the median and GM tend to coincide here as well, both at a fair distance from the arithmetic mean. Again, log transformation reshapes the data into a symmetrical pattern, in the same way as it did for the citation rates. It should be pointed out, however, that the category of Dutch books is deviating from the general pattern: here the arithmetic mean is closer to the median than the GM.

For the same reasons as those given earlier for citation rates, library impact and productivity ranking of individuals is best based on log *z*-scores. For group scores, again, there is greater freedom. Table 3 summarises, with respect to all indicators, for the five

Table 2 Median, arithmetic mean and GM for all indicators

Indicators		Arithmetic mean	Median	Geometric mean
1a	Citations (uncorrected)	27.3	13	12.3
1b	Citations (*PA*-corrected)	1.72	1.08	1
1c	Citations (per year)	1.04	0.58	0.57
2a	Library holdings: English, author	104.6	54.5	43.1
2b	Library holdings: English, editor	81.3	56	49.3
2c	Library holdings: Dutch	19.5	18	14.3
2d	Library holdings: Total	52.3	25	23.9
3	Productivity (pages authored per year)	115.3	74	79.7

Each of indicators 1a, 1b, and 1c presents citation rates over authors, but from different viewpoints: 1a is based on the original citations counts per author; 1b corrects for *PA* by using exponential regression; 1c presents the average citations per year per author

Table 3 Indicator statistics for disciplines, including standardized group scores

	Linguistics	Literature	History	Art history	Anthropology	Total
Citations						
Number of authors	59	84	72	27	22	264
Mean professional age	24.1	22.4	24.3	22.8	23.7	23.4
Total	2,101	1,313	2,597	359	846	7,216
Mean	35.6	15.6	36.1	13.3	38.5	27.3
Median	15	9.5	22	8	20.5	13
Geom. mean	14.8	8.3	18.6	6.6	19.5	12.3
Stdev.	58.1	20.4	41.7	15.6	55.4	41.5
Skewness	3.5	2.4	2.1	1.8	3.3	3.7
Standardized group scores						
Non-log (no *PA*-corr.)	0.20	−0.28	0.21	−0.34	0.27	0
Log (no *PA*-corr.)	0.14	−0.30	0.31	−0.48	0.35	0
Non-log: *PA*-corr.	0.19	−0.25	0.18	−0.34	0.28	0
Log: *PA*-corr.	0.12	−0.28	0.31	−0.54	0.41	0
Library holdings						
Number of authors	29	14	23	7	7	80
Titles	311	208	469	78	69	1,135
Total library holdings	11,892	9,897	29,627	2,838	5,132	59,386
Mean	38.2	47.6	63.2	36.4	74.4	52.3
Median	23	21	30	17.5	29	25
Geom. mean	20.3	21.5	27.5	19.0	35.0	23.9
Stdev.	44.0	80.5	111.6	46.6	83.1	87.0
Skewness	2.3	3.4	4.9	2.3	1.2	5.1
Standardized group scores						
Non-log	−0.16	−0.05	0.12	−0.18	0.25	0
Log	−0.13	−0.08	0.11	−0.18	0.30	0
Library holdings English, Author						
Number of authors	26	13	21	6	7	73
Titles	92	37	83	14	18	244
Total library holdings	4,864	5,072	12,374	1,019	2,195	25,524
Mean	52.9	137.1	149.1	72.8	121.9	104.6
Median	33.5	71	89	49.5	104	54.5
Geom. mean	27.4	61.1	55.9	38.6	70.7	43.1
Stdev.	53.9	144.2	205.7	70.1	97.2	146.0
Skewness	1.4	1.2	2.7	0.9	0.3	3.4
Standardized group scores						
Non-log	−0.35	0.22	0.30	−0.22	0.12	0
Log	−0.30	0.23	0.17	−0.07	0.33	0
Library holdings English, Editor						
Number of authors	20	10	16	4	6	56
Titles	89	34	103	18	25	269
Total library holdings	4,997	2,075	11,059	1,223	2,505	21,859
Mean	56.1	61.0	107.4	67.9	100.2	81.3

Table 3 continued

	Linguistics	Literature	History	Art history	Anthropology	Total
Median	52	42	73	54.5	89	56
Geom. mean	35.5	29.0	72.7	55.5	61.1	49.3
Stdev.	47.9	69.6	98.2	47.4	81.6	79.5
Skewness	2.0	1.9	2.3	1.7	0.8	2.5
Standardized group scores						
Non-log	−0.32	−0.25	0.33	−0.17	0.24	0
Log	−0.29	−0.46	0.34	0.10	0.19	0
Library holdings Dutch						
Number of authors	20	12	21	7	5	65
Titles	102	128	245	32	16	523
Total library holdings	1,509	2,437	5,587	453	207	10,193
Mean	14.8	19.0	22.8	14.2	12.9	19.5
Median	11	17	22	10.5	10.5	18
Geom. mean	10.6	14.5	16.9	10.8	10.4	14.3
Stdev.	12.0	12.1	14.1	9.9	8.3	13.3
Skewness	1.6	1.0	0.4	0.8	0.7	0.8
Standardized group scores						
Non-log	−0.35	−0.03	0.25	−0.40	−0.49	0
Log	−0.33	0.02	0.19	−0.31	−0.35	0
Productivity (pages)						
Number of authors	24	10	18	5	5	62
Total pages (annual)	2,503	876	3,118	173	480	7,150
Mean	104.3	87.6	173.2	34.6	96.0	115.3
Median	73.9	67.9	116.5	32.8	90.8	74.0
Geom. mean	75.9	70.2	113.6	31.9	90.9	79.7
Stdev.	80.9	61.3	169.6	15.2	35.6	113.4
Skewness	1.0	1.0	2.0	0.5	0.7	2.7
Standardized group scores						
Non-log	−0.10	−0.24	0.51	−0.71	−0.17	0
Log	−0.06	−0.15	0.40	−1.04	0.15	0

Statistics for the citations refer to citations per author, statistics for library analysis refer to library holdings per book title per author, and statistics for productivity refer to authored pages published per year per author

disciplines the standardized group scores, together with other statistics which can be helpful to interpret these indicators.

Figure 4 shows that group scores, when standardized, can easily be put together in one diagram. In our case the three main indicators show for the five disciplines quite similar profiles. The similarity indicates that the indicators are associated to some extent, which hints at a unifying structure underlying the different fields of citation impact, library impact and productivity. However, more data from different institutions will be needed to draw more firm conclusions. As regards the citation indicator extra caution is needed, while differences in amplitude between disciplines may be partly caused by citation index coverage (cf. Section "Shortcomings of standard citation analysis for the Humanities").

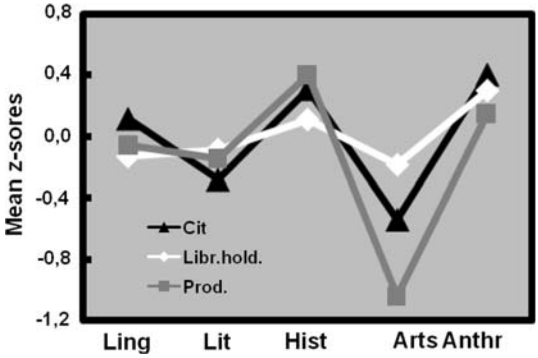

Fig. 4 Profiles of three main indicators (citations, library holdings, productivity) for five Leiden disciplines (based on log *z*-scores)

Table 4 Pearson correlations *r* of the indicators

	Cit	Cit (*PA* corr)	LH Total	LH Engl	LH Engl/ Author	LH Engl/ Editor	LH Dutch
Citations (*PA*-corr)	0.52**						
LH Total	0.29**	0.25*					
LH English	0.34**	0.24*	0.91**				
LH Engl/Author	0.39**	0.15	0.81**	0.92**			
LH Engl/Editor	0.26	0.12	0.72**	0.77**	0.38**		
LH Dutch	0.11	−0.10	0.11	0.20	0.23	0.16	
Productivity	0.29*	0.06	−0.02	0.04	0.04	0.07	0.27

LH library holdings, *PA* professional age. The correlations per cell are based on non-standardized non-log frequency data. Every cell correlation is derived from 47 to 78 researchers, depending on the data available

* Correlation is significant at the 0.05 level (2-tailed)

** Correlation is significant at the 0.01 level (2-tailed)

It is informative to compare the convergence of indicators at the discipline level, as shown in Fig. 4, with the Pearson correlations between indicators at the author level, as presented in Table 4. As could be expected, there are significantly strong correlations between the library holdings in total, the library holdings of authored books in English, and the library holdings of edited books in English; and also between citation rates with and without correction for *PA*.

The citation rates (both with and without *PA*-correction) show significantly strong correlations with the over-all library holding rates. However, if, with the library holdings, we make a distinction between books in English and books in Dutch, the strong correlation with citation rates appears to exist only for the books in English. The reason for this is the bias in favour of publications in English, which unites the citation data and the library holdings of libraries in the United States. A practical conclusion to be drawn from this is that the library holding indicator for books in Dutch (or, *mutatis mutandis*, in languages other than English) recommends itself as a valuable indicator, supplying some very specific information not found with the other indicators.

However, at the author level there seems to exist no clear correlation between the productivity rates and the library holding rates, something which at the higher aggregation level of the disciplines was apparent though (Fig. 4).

Discussion

There are a few points that require further discussion and examination. I will briefly point these out.

Feasibility

To test the full capability of the proposed assessment methods, a much larger amount of data (both cross-disciplinary and cross-institutional) will have to be assembled. To do this cost-effectively, automated procedures must be developed to diminish work-load.

The greatest bottle-neck in the gathering of citation data is author identification. For non-source publications the citation indexes are full of inaccuracies in the rendering of initials and surnames. Also the disentangling of homograph names is often fraught with difficulties.

Problems in the gathering of library holdings are: the imperfect matching in *WorldCat* of the title descriptions of different libraries; the distinction which has to be made between authored and edited books (which requires the inspection of the full title descriptions); and the coupling of re-editions and translations with the original works.

The presentation of rankings

Given the limitations of a purely metric impact and output analysis (perhaps in the Humanities even greater than elsewhere), and the fact that this may complicate acceptance, it is worthwhile to carefully consider the way in which the outcomes of the ranking operations are presented. Useful strategies may be the simultaneous presentation of all indicator outcomes to support a multifaceted assessment, and ranking within the boundaries of a discipline, so as not to make generalizations that do not hold.

Another aspect is the skewness of the data we are confronted with. Hence, in the top segment the distances between individuals and groups are always quite distinct and relevant, but in the low performing region the mutual distances are often so small, that they no longer are significant. Therefore, especially with individual rankings, it can in my opinion be more realistic to define score intervals with the help of quantiles. It may be more illuminating to describe that an individual belongs to a specific quartile or decile, than that he or she occupies say rank 248.

Conclusions and prospects

A primary concern throughout this study has been the skewness of our data. This incited us, when dealing with assessment indicators, to opt for the log data, and to do most ranking with the help of log z-scores.

Although the results for the different indicators showed a certain convergence, there is a need for profiles that present all indicator outcomes in parallel. For instance, only by zooming in on library holdings of Dutch books, we counterbalance the Anglo-American bias of most other indicators.

Over-all it appeared that Humanities research lends itself to bibliometric analysis no less than the other sciences. However, the analytic methods have to be adjusted to the specifics of the Humanities and to availability of current data. For assessment indicators to be representative, fair, and relevant, we had to stretch the use of available citation data beyond what is usually done in standard citation analysis. Library holding indicators could mend some of the most apparent deficiencies of citation analysis. Given the importance of the Internet, indicators from that area are bound to follow. Particularly the citation counts for books supplied by Google Scholar will be of great value.

In the future, more data will be needed from different institutions in order to judge local performances against the background of what is internationally viable for a discipline.

In another study I hope to demonstrate that the data is obeying some specific, mathematically definable bibliometric laws, when seen from the view-point of distribution and growth.

Acknowledgments I wish to express my gratitude to the Executive Board of Leiden University, and especially its former Vice-Rector magnificus Professor Ton van Haaften, for the opportunity given to carry out this study. I am indebted to Professors Geert Booij and Wim van der Doel, Deans of the Leiden Faculty of Humanities and their staff, and Piet van Slooten, Director of Academic Affairs at Leiden University, and his staff for their encouragement and support. The project would not have been possible without Professor Anthony van Raan, Director of CWTS, who offered the stimulating environment of his institute and who read the manuscript. Henk Moed, Ton Nederhof, Martijn Visser, and my other colleagues at the CWTS helped me by commenting on parts of the preliminary report and by supplying extra data. I am grateful to Henk Moed for his encouraging me to investigate library catalogues as a bibliometric source. I thank the peer reviewers for their helpful comments.

Appendix: mathematics of *PA*-correction for citations

Starting from the data as it is plotted in the scatter diagram of citation rates per author over *PA* (Fig. 2, main text), we can *PA*-correct the citation rates by using the exponential regression:

$$ce(y_k) = \frac{y_k}{E(y_k)} = \frac{y_k}{\alpha e^{\beta x_k}}, \tag{1a}$$

where $ce(y_k)$ denotes the exponentially corrected citation score of author y with professional age k. The lower bound of all points $ce(y_k)$ is 0, and their GM is $\bar{y}_{\mathrm{GM}} / \alpha e^{\beta \bar{x}} = 1$. The Cartesian point $\left(\bar{x}, \alpha e^{\beta \bar{x}}\right) = (\bar{x}, \bar{y}_{GM})$ we call the geometric centre of gravity.

Alternatively, *PA*-correction can be based on the linear regression (*cl* standing for linear correction):

$$cl(y_k) = y_k - E(y_k) = y_k - (\gamma + \delta x_k). \tag{2a}$$

The arithmetic mean of all scores $cl(y_k)$ is $\bar{y} - (\bar{y} + \delta \bar{x}) = 0$. The centre of gravity in the (x, y)-plane is the Cartesian point $(\bar{x}, \gamma + \delta \bar{x}) = (\bar{x}, \bar{y})$.

Since there are advantages in working with linear data, we may translate (1a) into linear form by taking logs. Because $\log[ce(y_k)] = cl(\log y_k)$, we thus apply the following log transformation of (1a):

$$cl(\log y_k) = \log y_k - (\log \alpha + \beta x_k). \tag{1b}$$

The arithmetic mean of all $cl(\log y_k)$ is $\overline{\log y} - (\log \alpha + \beta \bar{x}) = 0$.

So far we have two linearly corrected scores to work with, (1b) and (2a), which, if we take their origin into account, are fundamentally different in spite of their external similarity, since (1b) is indirectly based on exponential correction of the original counts, while (2a) is directly based on linear correction of the same counts. Hence, both (1a) and (1b) can be seen as representing a model based on exponential correction, while (2a) provides a model based on linear correction.

Finally we standardize scores (1b) and (2a), so that they not only have mean 0, but also standard deviation 1. This is done by applying the standardizing operation $z = (x - \bar{x})/s$, where s is the standard deviation of all x. It should be reminded that a different type of standardizing procedure was applied above with regard to (1a). There we calculated the ratio $x/\bar{x}$ with mean 1 and lower bound 0, without standardizing the deviation.

Consequently, the standardized *PA*-corrected scores for the exponential and the linear model, corresponding with (1b) and (2a), are respectively:

$$\log z_k = \frac{\log y_k - (\log \alpha + \beta x_k)}{s}, \tag{1c}$$

and

$$z_k = \frac{y_k - (\gamma + \delta x_k)}{s}. \tag{2b}$$

We can add the following standardized scores without *PA*-correction:

$$z = \frac{y - \bar{y}}{s}, \tag{3a}$$

and

$$\log z = \frac{\log y - \overline{\log y}}{s}. \tag{3b}$$

Group means (g standing for group) $\bar{z}_{kg}, \overline{\log z_{kg}}, \bar{z}_g, \overline{\log z_g}$ (in the main text denoted by:$\bar{z}_x$, $\overline{\log z_x}$, $\bar{z}$, $\overline{\log z}$) are obtained by calculating the averages of the individual scores (2b), (1c), (3a) and (3b), respectively of all members of the group in question. It should be noted that the parameters α, β, δ, γ, the means $\bar{y}$, and $\overline{\log y}$, and the standard deviation s are derived from the reference group. In our case as reference group we always have the union of the groups (samples) for which the group means are computed.

The sampling distributions of the aforementioned group means are normal distributions with $\mu = 0$ (CLT). Should we use, instead, the aforementioned geometric scores of (1a), then the sampling distribution is a lognormal distribution ($\mu = 1$).

References

Archambault, É., et al. (2006). Benchmarking scientific output in the social sciences and humanities: The limits of existing databases. *Scientometrics,68*, 329–342.

Burrell, Q. L. (1990). Empirical prediction of library circulations based on negative binomial processes. In L. Egghe & R. Rousseau (Eds.), *Informetrics 89/90* (pp. 57–64). Amsterdam: Elsevier Science Publishers.

Butler, L., & Visser, M. S. (2006). Extending citation analysis to non-source items. *Scientometrics,66*, 327–343.

Cronin, B., & Snyder, H. (1997). Comparative citation rankings of authors in monographic and journal literature: A study of sociology. *Journal of Documentation,53*, 263–273.

Dekking, F. M., Kraaikamp, C., Lopuhaä, H. P., & Meester, L. E. (2005). *A modern introduction to probability and statistics: Understanding why and how*. London: Springer.

de Solla Price, D. J. (1970). Citation measures of hard science, soft science, technology, and nonscience. In C. E. Nelson & D. K. Pollock (Eds.), *Communication among scientists and engineers* (pp. 3–22). Lexington: Heath.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. arXiv:physics/0508025

Larivière, V., et al. (2006). The place of serials in referencing practices: Comparing natural sciences and engineering with social sciences and humanities. *Journal of the American Society for Information Science and Technology,57*, 997–1004.

Leydesdorff, L., & Bensman, S. (2006). Classification and powerlaws: The logarithmic transformation. *Journal of the American Society for Information Science and Technology,57*, 1470–1486.

Lindholm-Romantschuk, Y., & Warner, J. (1996). The role of monographs in scholarly communication: An empirical study of Philosophy, Sociology and Economics. *Journal of Documentation,52*, 389–404.

Line, M. B. (1979). The influence of the type of sources used on the results of citation analyses. *Journal of Documentation,35*, 265–284.

Linmans, A. J. M. (2008). *Wetenschappelijk onderzoek in de Faculteit der Letteren van de Universiteit Leiden: Een onderzoek naar bibliometrische indicatoren voor het bepalen van impact van wetenschappelijk onderzoek in de geesteswetenschappen*. Leiden: CWTS.

Luwel, M., Moed, H. F., Nederhof, A. J., et al. (1999). *Towards indicators of research performance in the social sciences and humanities: An exploratory study in the fields of Law and Linguistics at Flemish Universities*. Brussel: Vlaamse Interuniversitaire Raad.

Moed, H. F. (2005). *Citation analysis in research evaluation*. Dordrecht: Springer.

Nederhof, A. J. (2006). Bibliometric monitoring of research performance in the Social Sciences and the Humanities: A review. *Scientometrics,66*, 81–100.

Snow, C. P. (1959/1964). *The two cultures: And a second look*. Cambridge: Cambridge University Press.

Torres-Salinas, D., & Moed, H. F. (2009). Library catalog analysis as a tool in studies of social sciences and humanities: An exploratory study of published book titles in Economics. *Journal of Informetrics, 3*, 9–26.

van Leeuwen, Th. N., Moed, H. F., Tijssen, R. J. W., Visser, M. S., & van Raan, A. F. J. (2001). Language biases in the coverage of the Science Citation Index and its consequences for international comparisons of national research performance. *Scientometrics,51*, 335–346.

van Raan, A. F. J. (2006). Statistical properties of bibliometric indicators: Research group indicator distributions and correlations. *Journal of the American Society for Information Science and Technology,57*, 408–430.