Diem, Andrea; Wolter, Stefan C.

The use of bibliometrics to measure research performance in education sciences

Aarau : Swiss Coordination Centre for Research in Education (SCCRE) 2011, 28 S. - (Leading House Working Paper; 66)

Working Paper No. 66

The Use of Bibliometrics to Measure Research Performance in Education Sciences

Andrea Diem and Stefan C. Wolter

Universität Zürich
IBW – Institut für Betriebswirtschaftslehre

b
UNIVERSITÄT
BERN

Leading House Working Paper No. 66

The Use of Bibliometrics to Measure Research Performance in Education Sciences

Andrea Diem and Stefan C. Wolter

September 2011

The Swiss Leading House on Economics of Education, Firm Behavior and Training Policies is a Research Programme of the Swiss Federal Office for Professional Education and Technology (OPET).

www.economics-of-education.ch

The Use of Bibliometrics to Measure Research Performance in Education Sciences

Andrea Diem* / Stefan C. Wolter**

September 2011

Abstract

This paper uses bibliometric data to investigate the research performance of Swiss professors in the field of education sciences. The analyses are based on two separate databases: Web of Science and Google Scholar. A comparison of the various indicators used to measure research performance (quantity of publications and citation impact) from the two data sources indicates highly positive correlations between all of them, to a greater or lesser degree. At the same time, there is evidence that significant individual factors that would serve to explain the great variance in research performance can be identified only if the Web of Science is used as a benchmark of research performance. However, the Web of Science inclusion policy is associated with certain issues that put some research authors at a disadvantage. Therefore, problems currently exist in regard to both citation databases when used to benchmark individual research performance: Web of Science adopts a selective approach, but some of the criteria employed are problematic. Google Scholar on the other hand is so inclusive that it is virtually impossible to identify explanatory variables for the existing major individual differences in research performance.

Keywords: bibliometrics, education sciences, research performance, scientometric methods, science research

JEL-Codes: I23, I29

* Swiss Coordination Centre for Research in Education (SCCRE), Aarau
** Swiss Coordination Centre for Research in Education (SCCRE), Aarau, and University of Bern, CESifo and IZA

Communicating author: Andrea Diem (andrea.diem@skbf-csre.ch)

1. Introduction[1]

The practice of rating research performance on the basis of bibliometrics is ubiquitous in academic research. Rating has systematic value in providing a rationale to justify the allocation of research funds and for quality assurance in research programs and projects, in that way enabling strategic planning at universities, and it is also used for international benchmarking of universities and faculties in an increasingly competitive battle for scientific and economic resources (see European Commission 2010). Bibliometric data is also used as means of rating individual researchers, as a basis both for allocating funds at universities and research institutes, and influencing researchers' job-seeking chances. In view of the growing importance of bibliometric data for systematic and individual rating of research performance, various international and national initiatives to develop new methods and approaches for measuring research outcomes (examples being the EERQI project[2] on a European level or the initiatives of CRUS[3] for Switzerland) are ongoing.

The popularity of bibliometrics probably resides in the fact that the information is highly compact, easy to handle, and likely to be objective.[4] Nonetheless, bibliometric methods have their critics among experts (and, naturally, among researchers as well). The criticisms concern general methodological challenges that call into question the appropriateness of the measuring instrument per se (see Adler, Ewing, and Taylor 2009, Fröhlich 1999, Jokić and Ball 2006, Moed 2005, Neuhaus 2010). In particular as far as the humanities and social sciences are concerned (and other disciplines such as law), there are additional difficulties involved in measuring the quality of research performance, as publication and citation practices in these areas differ significantly from customary practice in other academic disciplines, which casts even more doubt on the validity, and fitness for purpose, of bibliometric evaluation (see Hicks 1999, Moed 2005, Nederhof 2006). As a result, comparisons of research performance across different disciplines and, in some cases, even within one and the same discipline (regional language disparity, differences between sub-disciplines, cohort effects), always need to be analysed in light of the differences in the importance and evaluation of research performance by bibliometric techniques.

This paper is therefore limited to investigating the fitness-for-purpose and soundness of bibliometric parameters in measuring and elucidating the research performance of individual researchers in one specific research area (education sciences) and one country (Switzerland). Few studies are available so far on the (bibliometric) measurement of research performance in education sciences (examples include Botte 2007, Budd 1988, 1990, Budd and Magnuson 2010, Davarpanah 2009, Dees 2008, Earp 2010, Fernández-Cano and Bueno 1999, Hornbostel and Keiner 2002, Keiner 1999, Klusmeyer, Reinisch, and Söll 2011, Kroc 1984, Shin 2004, Silverman 1985, Smart 1983, Smart and Elton 1981, Smart and McLaughlin 1982, Tight 2008, Togia and Tsigilis 2006, van Aalst 2010), and the results of existing studies may not apply "as is" to other countries and periods. Furthermore, it is worth bearing in mind that publishing practice in education sciences has traditionally been of a nature that hardly encourages the use of bibliometric methods to measure research performance. However, there has been a definite trend in recent years toward adopting the practice of publishing in peer-review journals. In addition, databases such as Google Scholar that cover a wide range of publications now

[1] The authors thank Daniel Munich and Olivier Rey for helpful comments on the manuscript.
[2] www.eerqi.eu
[3] www.crus.ch/die-crus/koordiniert-harmonisiert/projekt-mesurer-les-performances-de-la-recherche.html
[4] For an introduction to bibliometrics in science communication, see Jokić and Ball (2006).

exist, with the result that the use of bibliometrics is now a feasible option in education sciences as in other areas.

This paper investigates three specific issues:

Firstly, two different databases are used to measure individual research performance: the more restrictive Web of Science, and the more extensive Google Scholar database. The use of two different databases with very different inclusion criteria for research performance is intended as a means of finding out how much the rating for individual research performance depends on the database used as the source of the bibliometric information.

Secondly, the two databases are used to construct quantitative and qualitative measures of individual research performance. The number of publications contained in the databases provides a quantitative measure of individual research output, and the citation count provides a qualitative measure[5], i.e., the citations actually denote a research outcome, namely the impact of the published research papers on other people's research. This second step investigates the connections between output and outcome and whether these connections depend on the database used for the comparison.

Thirdly, we attempt to explain inter-individual differences in research output and outcome on the basis of individual and institutional characteristics of the researchers. The main question here is whether different research performance levels can be explained by observable characteristics of the researchers involved, and if so, which ones.

The paper is structured as follows: chapter 2 introduces bibliometric indicators, production of publications in educational research, and explanatory factors for research performance. Chapter 3 presents our database and the basics of the method. Chapter 4 gives a descriptive summary and outlines the connections between different research performance indicators. Chapter 5 presents the findings on the explanatory factors for research performance. Finally, chapter 6 summarizes the insights obtained.

2. Bibliometric indicators, science communication in education sciences, and explanatory factors pertinent to research performance

2.1. Indicators for measurement of research performance

Bibliometric indicators used to measure research performance are mainly based on two central elements: number of publications and citation count. The statistics based on the number of publications primarily reflect the quantitative output of research activity. In contrast, there is little agreement on what the figures based on citations exactly measure, as the reasons for citing a paper may be highly disparate (see Jokić and Ball 2006, Krampen et al. 2007, Moed 2005).[6] This paper

[5] Because Web of Science only includes publications that have been published in scientific journals that are listed in the Social Sciences Citation Index, even a purely quantitative analysis of these publications implies a qualitative element, as the vast majority of included publications will have been subjected to review prior to acceptance for publication.

[6] Citations may be meaningless or have negative connotations and citation impacts may be inflated by "citation cartels" and self-cites.

interprets the individual citation impact more as a measure of the response elicited by a piece of research in the academic community. Another form of citation impact is one in which scientific journals have an impact factor which, in turn, is based on the frequency of citation of articles published in the particular journal. Journal-based impact factors of this kind are relatively widespread, although not uncontroversial (Schulze, Warning, and Wiermann 2008), and are used to attach a weighting to each article published by a researcher, based on the impact factor of the journal in which the material was published. This does not measure the impact of the actual article or researcher, but does deliver a qualitative statement about the article, as it can be assumed that the standards of a journal with a high impact factor will be superior, in that it is more difficult to be accepted for publication in that kind of journal.

Extensive journal rankings exist for many disciplines, which makes it possible to weight the (quantity of) published papers by journal quality. Attempts have been undertaken to draw up similar journal rankings in educational research. Currently available lists do not adequately cover the journals served by Swiss researchers (ERA Journal Ranking), do not indicate journal quality based on classification (European Reference Index for the Humanities (ERIH) of the European Science Foundation), fluctuate significantly between the various sub-disciplines (Budd and Magnuson 2010, Earp 2010, Fairbairn et al. 2009, Togia and Tsigilis 2006) or are criticized on grounds of poor overall validity (Corby 2003, Haddow and Genoni 2010, Luce and Johnson 1978, Rey 2009, Smart 1983, Wellington and Torgerson 2005). In consequence, there seems to be little benefit at present in using journal rankings in educational research.

In addition to basic statistics on publication and citation frequency, bibliometric analysis uses numerous other indicators that differ in terms of weighting on aspects, such as skewed distribution of publication and citation frequency (see Bornmann, Mutz, and Daniel 2008, Panaretos and Malesios 2009, Todeschini 2011), differences in the quality of the publications, number of co-authors (Rauber and Ursprung 2008), length of publication, and differences in communication cultures between disciplines (see De Witte and Rogge 2010, Hofmeister and Ursprung 2008, Nederhof, Luwel, and Moed 2001, Prathap 2011, Ritzberger 2008, Schulze, Warning, and Wiermann 2008).

2.2. Brief outline of science communication in education sciences

Very little is known about science communication at Swiss institutes of education sciences so far. Precise figures on current publication activity are not available. Although the individual universities can be assumed to document research performance in the form of periodic publication lists, these lists are not always available to the public, or may be available only for a limited period, or may not be made available in a standardized form that would enable comparison between individual researchers. A comparatively old study by Cusin, Grossenbacher, and Vögeli-Mantovani (2000) provides a number of pointers on publication output. This study is based on detailed investigation of the publication frequency of the education sciences departments of three universities (Zurich, Freiburg (German speaking department) and Geneva) in the 1996-1998 period. Classification by type of publication revealed the following pattern: of the approximately 1,100 publications studied, book chapters (23%) and articles in user-oriented periodicals (24%) each accounted for just under one-quarter each, while articles in scientific journals accounted for about one-fifth (21%). Monographs and conference papers each accounted for seven percent (other publications: 18%).[7] However, there were significant differences in the relevant percentages between the individual departments. A look at the provenience of the media in which the scientific articles were published indicates a heavy local bias. About two-thirds of articles were published in a Swiss journal or journal published in the same language as the university department in question. Only 17 percent of the articles appeared in Anglophone journals.[8] Furthermore, analysis of the scientific articles (n=234) indicated a heavy focus on just a few (national) outlets.

Studies conducted in German-speaking areas (Dees 2008, Keiner 1999) reveal definite disparity in terms of selected type of publication, publication outlets, and publication language between different educational research departments. Disparate publication practices have also been observed between the various sub-divisions of education sciences (Keiner 1999). Similar findings are also reported from non-German-speaking areas (Silverman 1985, van Aalst 2010).

Education sciences papers very often have a single author (Dees 2008, Hornbostel and Keiner 2002, Keiner 1999, Klusmeyer, Reinisch, and Söll 2011). More than half of the publications in the Dees study (57%) had a single author, and one-quarter (25%) had two authors. The average number of authors was 1.8; Keiner (1999) puts the figure at 1.1 to 1.2. Authorship tends to be larger for papers published in English (Dees 2008).

Findings available to date on distribution of publication output and outcome show major variation in research performance in education sciences, as in other disciplines (see e.g. Aaltojärvi et al. 2008, Bernauer and Gilardi 2010, Rauber and Ursprung 2008), both between researchers and between different research departments. No publications were identified in the Education Information System database for one-third of the education sciences professors in Germany during the 1997-1999 period. One solitary publication was identified for another 18 percent (Hornbostel and Keiner 2002). This skewed distribution in terms of publication and citation frequency can only be explained in part by variations in coverage of educational research literature in the individual sub-areas (Corby 2001) or by differences in citation practices (Kroc 1984).[9]

[7] Other figures from German-speaking countries reflect the following: the more up to date percentages in the Dees study (Dees 2008) with a somewhat different operationalization indicated 33% journal articles, 47% book chapters, 15% books and 5% other publications. A study on the basis of data from the Education Information System (Fachinformationssystem Bildung) (Hornbostel and Keiner 2002), which covers publications from Germany, Austria and Switzerland calculated the following percentages for professors: 59% journal articles, 25% book chapters, 16% monographs; these figures do not include editorships, basic didactics, grey literature, etc. New figures for the area of vocational and business education which are based on the same database show a composition of 40% journal contributions, 40% book chapters and 20% monographs (Klusmeyer, Reinisch, and Söll 2011).

[8] A heavy national focus / use of the national language was also observed in the German study by Dees (2008): 88 percent of the publications analysed were written in German. In the study limited to the area of vocational and business education (Klusmeyer, Reinisch, and Söll 2011) non-German publications amounted to 3.5% only.

[9] Investigations in related social science research areas suggest that the skewed distribution for research performance is not solely explained by the fact that researchers differ in the types of publication they prefer and are more or less likely to be included in databases on that account. Researchers with a high level of publishing activity in one particular type of publication (monograph, book chapter, journal article) tend to have higher publishing outputs in respect of other types of publication as well (Puuska 2010). Nor is the skewed distribution likely to be due to a quantity versus quality trade-off; for instance, a study by Bernauer and Gilardi (2010) looking at political science shows that researchers who publish more articles also tend to have higher rates of publication in journals with a higher impact factor.

2.3. Determinants of individual research performance

The variance in research activity can be explained by a number of individual and institutional characteristics, as evidenced by the results of studies in various disciplines. There follows a brief outline of empirical findings on the three potential main individual determinants amenable to analysis in this empirical review:

Age: From an empirical point of view, results vary widely in terms of a possible correlation between age and research performance. Some are confirmatory and some are contradictory. The “life cycle model” based on human capital theory, which models scientific production as a function of investments and human capital depreciation, postulates an inverted U-shaped correlation between (academic) age and research output. Empirical analysis mostly indicates a trend in keeping with this theory, according to which publication activity increases during the first years in academia and then gradually plateaus (Gonzalez-Brambila and Veloso 2007, Rauber and Ursprung 2008, van Ours 2009) or falls off (albeit with linear specifications of age; see Carayol and Matt 2006, Levin and Stephan 1991, Smeby and Try 2005). Alongside the human capital theory and depreciation over time of the human capital that drives research output, alternative explanations exist for the initial rise in the research output curve, followed by a plateau or actual decline. For instance, organizational and administrative duties at a university are likely to increase with age, leaving less time for research and hence for publishing (see Knorr et al. 1979). Alternatively, incentive structures may change (acceptance of senior academic positions such as dean or head), or there may be less incentive to do research (tenured position).

In addition to individual age effects, cohort effects are likely to apply (see Hall, Mairesse, and Turner 2005) which tend to skew analysis of age effects in purely cross-sectional studies. Both the number of publications and citation count have been on the rise in recent decades (Gonzalez-Brambila and Veloso 2007, Graber, Launov, and Wälde 2008, Moed 2005, Rauber and Ursprung 2008). In cross-sectional studies, this results in a tendency to underrate the activity of older researchers. To enable analysis of output as a function of academic age, the outputs of identical cohorts would therefore have to be compared over different periods of time.[10]

Another explanation for the age effect is the observed increase in multiple authorship by age (Moed 2005); correction of publication output by the number of co-authors would indicate a much lower growth rate. However, given the low overall rate of multiple authorship in educational research, the associated bias is likely to be negligible.

An interesting feature and one that is probably specific to the field of educational research is that there is not much of a correlation between biological age and research age (years since Ph.D.). This is because the professional biographies of professors in the field of education sciences display a high level of heterogeneity. As a result, the effect of both age variables can be tested together. A distinction of this kind would not be possible in most other scientific disciplines as there is usually a strongly positive correlation between the two variables.

Professional category (position in the academic hierarchy): A number of explanations can be posited for the correlation between professional category and individual research performance. There are

[10] In theory, this form of analysis would also be feasible with the data in this paper. In practice, the low overall number of observable professors is a prohibitive factor.

two possible explanations for positive correlations. The first is what is known as a selection effect. "Good" research scientists in particular, i.e., those with successful research and publishing histories, are more likely to be promoted to higher positions. This produces a positive correlation with a seemingly causal relationship between publishing activity and likelihood of being in a higher professional category. The second explanation posits causality in the opposite direction, i.e., that the higher position has a positive impact on research performance, as the more prestigious position is more likely to be associated with favourable conditions for research work (number of assistants, access to research projects, less time taken up with teaching). Other theories posit a negative causal relationship: the incentive to do heavy-duty research may diminish as soon as the person has achieved the goal of acquiring a tenured position. However, the findings of the empirical studies conducted to date indicate an unmistakably positive correlation between professional category and research productivity (see Aaltojärvi et al. 2008, Carayol and Matt 2006, Puuska 2010, Smeby and Try 2005), but do not give a clear indication of the direction of causality.

Gender: Various studies show that the publishing output of female researchers is lower than that of their male counterparts (Aaltojärvi et al. 2008, Kyvik 1996, Larivière et al. 2011, Puuska 2010, Rauber and Ursprung 2008, Smeby and Try 2005). In contrast, findings indicating positive effects of female gender on research output are rare (De Witte and Rogge 2010). Possible explanations for the negative correlation might be poorer integration of women in the research community and lower levels of support for female researchers (smaller networks, less women in influential positions such as editorial boards, etc.). Another hypothesis is that female researchers and professors have less time to devote to research owing to family commitments and therefore do in fact produce less research output.

Alongside the individual determinants of individual research performance, institutional factors explain some of the variance in research output. Identified factors include peer effects, i.e., the research performance of colleagues (Carayol and Matt 2006), age structure, i.e., more experienced or less experienced colleagues (Bonaccorsi and Daraio 2003), and department size (see Carayol and Matt 2006). As far as these factors are concerned, the expectation is that researchers working in an environment with other productive researchers will produce more research themselves (positive spill-over effects). However, it is always relatively difficult to furnish empirical evidence of a causal relationship between a researcher's output and that of his or her colleagues, because an alternative explanation for any such correlation would always be that a department with a productive research environment is more likely to recruit productive new researchers; this would be a selection effect, pure and simple.

3. Database and methods

3.1. Basic population and data sources

Our study population is made up of all professors (male and female) in the field of education sciences[11] employed at a Swiss university on the index date (10 September 2010). The population includes all tenured professors, titular professors, associate professors and assistant professors. It

[11] Professors of education sciences are defined as those individuals classified in the CRUS directory (www.proff.ch; updated February 2010) in the education sciences field.

does not include honorary professors, outside lecturers ("Privatdozierende"), emeritus professors, guest professors or visiting professors. Each professor is a unit of analysis (N=51).

We investigate our research questions using two different research performance data sets obtained on the basis of a top-down approach: the Thomson Reuters Web of Science citation database ([v.5.1], see www.webofknowledge.com) and Google Scholar (using Publish or Perish software[12]). The following paragraphs provide a brief outline of the two data sources and implications for bibliometric analysis.

The Thomson Reuters *Web of Science* is probably the most popular citation database for calculating bibliometric statistics.[13] The database covers high-quality scientific publications (peer-reviewed journal articles in most cases). The main emphasis has traditionally been on the exact sciences and natural sciences (including medicine). An extensive humanities and social science database has been established in the meantime. 2,257 journals are included at present in the Social Sciences Citation Index (SSCI). A total of 213 journals are indexed in the education sciences categories (namely, "education & educational research", "education, special" and "psychology, educational"). In view of the quantitative significance of educational research in academic research as a whole, it appears that there is no under-representation of educations sciences, at least in terms of the number of indexed journals in the SSCI.

The criteria for including publications in the database are heavily based on customary practice in the natural sciences. Monographs and book chapters, which are the main publishing outlets for many of those engaged in humanities and social science research, are not included in the Web of Science, resulting in low overall coverage of education sciences literature in all of its published forms (Corby 2001, Togia and Tsigilis 2006). Likewise, non-English-language and non-international journals have a lower chance of inclusion in the database (see Archambault et al. 2006, Nederhof 2006, Van Leeuwen 2006); three-quarters of the education sciences journals are published by US or UK publishers. Since many education sciences papers are contextualized in a local setting, and most are intended solely for the local research community, their chances of being included among the indexed journals are remote. The 2000 study by Cusin, Grossenbacher, and Vögeli-Mantovani confirms this expectation for Switzerland during the years 1996-1998. Only 7.4 percent of the articles published in journals at the three university departments investigated were indexed in the SSCI. The Dees study conducted ten years later (2008) estimated a figure of 14% for publications from 14 German departments of education sciences, again indicating a low proportion of coverage.

The alternative (or complementary) *Google Scholar* citation database does away with most of the deficiencies of the Web of Science (see Harzing and van der Wal 2008, Jacsó 2008, Meho and Yang 2007), but has other drawbacks of its own. The scientific literature in Google Scholar has a much broader base. Importantly, it takes account of all of the different types of publication: journal articles, monographs, book chapters, conference contributions, reports and grey literature. Another major advantage of the database is that it embraces a broad range of non-English-speaking literature. However, unlike with Web of Science, inclusion of publications is low-threshold and not subject to quality control. The disadvantage is that literature can enter the database that does not

[12] Harzing, A.W. [2010] *Publish or Perish,* version 3.1.3910 (www.harzing.com/pop.htm).
[13] For an extensive account of the Web of Science citation database, see Jokić and Ball (2006).

comply with established scientific criteria.[14] Poor data base quality is evident in the citations (García-Pérez 2010, Jacsó 2008).

Van Aalst (2010), who compared various databases on the basis of citation impacts for three areas of education sciences, concluded that, despite its weak points, Google Scholar nonetheless delivers valuable bibliometric information.

The two databases differ significantly from each other as regards degree of coverage, type of publications included, and data quality / homogeneity. However, combining the two data sources provides the opportunity to conduct comparative analyses and hence to establish the effect that the use of a specific bibliometric data source has on the results of analysis.

The bibliometric data of the individual researchers was acquired in the period from 29-30 September 2010 (Web of Science) and 15-22 October 2010 (Google Scholar). All publications and citations in the Thomson Reuters database were identified that were indexed in the Social Sciences Citation Index (SSCI) or Arts & Humanities Citation Index (A&H-CI).[15] The Citation Report delivered information on number of publications, citation count, citation count excluding self-cites, and *h*-index.

Every hit in Google Scholar was included that was listed in *Publish or Perish* under the heading "Social Sciences, Arts, Humanities". The software provides information on number of publications, citation count, citation count per author, *h*-index and number of authors. The acquired data underwent rigorous cleansing. Duplicate titles were eliminated, as were titles that were not specifically linked to an actual piece of research[16]. In addition to conventional publication types, the cleansed database also contains conference papers (not included in the SSCI) and hits with references to review articles. All these particulars were identified for the entire period.

3.2. Dependent variables: research performance indicators

Our analysis is conducted using a number of different indicators, which allows us to compare a variety of bibliometric statistics and correlations. Our method takes account of indicators that are primarily intended to measure the quantitative dimension of research output (number of publications) and indicators that reflect the level of response elicited in the academic community (outcomes) (in this instance: citation impact). Data for all indicators was accumulated for each individual scholar's entire life's work (see Linmans 2010). To enable specific sub-analysis of age effects, only the accumulated publications per researcher within the most recent time slot (2005-2010) were used. Our main dependent variables are:

- *Number of publications*: This variable includes all publications throughout the person's productive period. For the variable established on the basis of Web of Science, only the number of journal publications was taken into account (i.e., no editorials, reviews, etc.).
- *Number of publications 2005-2010*: This variable is based on the above variable (number of publications) but only includes literature published in the 2005 to 2010 period.

[14] It is possible that a significant proportion of the publications "only" constitutes material addressed to users in the field rather than research results in the strict sense. For instance, 16% of documents contained in the Education Information System have been shown to be teaching materials or basic didactic information (Hornbostel and Keiner 2002).
[15] Conference papers (CPCI-SSH) were not included.
[16] For instance references to publishers or university homepages.

- *Citation count*: This variable covers all citations documented in the databases for the indexed publications. In respect of the Web of Science-based variable, only citations referring to an actual article were used. Self-cites were excluded.
- *Citation count per publication*: this variable is the ratio of citation count to number of publications.
- *h*-index: a researcher with an index of *h* has published *h* papers, each of which has been cited by others at least *h* times.

The data was not corrected for number of co-authors[17] or article length.

3.3. Explanatory variables

To explain variances in research performance, the effect of the following factors will be looked at in greater depth:

- *Academic age*: This variable measures the number of years (in 2010) since obtaining a doctorate. Squared terms are also inserted into the analysis to investigate non-linear correlations.
- *Biological age*: This variable gives age in 2010. Again, squared terms are taken into account.
- *Professional category*: This variable is operationalized as a dummy variable. It assumes a value of 1 if a person has a tenured professorship and 0 for all other cases.
- *Gender*: This dummy variable assumes a value of 1 if the subject is a woman.
- *Language region*: This variable is operationalized as a dummy variable. Researchers working at a French-speaking department assume a value of 1.

To avoid bias attributable to inter-department differences, our analysis includes a number of control variables. We use dummy variables for the various departments, dummy variables for the different areas of studies (i.e., didactics, general pedagogy, adult education, sociology/systems research, child psychology/special needs teaching/anthropology), the number of fellow-professors in the department, and the average productivity of colleagues in the department (mean number of publications, mean citation count, etc.).

3.4. Analysis methods[18]

Multivariate regressions for analysis on the basis of Google Scholar data was mainly done using OLS regression with logarithmized dependent variables. Ordered probit models were calculated to verify the results. In respect of Web of Science variables, logarithmization of the variables did not achieve normal distribution. Two-stage estimator models were therefore used (negative binomial logit hurdle models or Poisson logit hurdle models; see Winkelmann 2008). Hurdle models are a highly suitable method of analysis for questions of this kind because, in the presence of a large number of observations with 0 values, the factors that can explain who publishes at all, and the factors that can

[17] Our figures show the following findings in respect of number of authors: the mean (average) number of authors per professor is 2.2. This authorship figure is somewhat higher than reported by Dees (2008) and in the older Keiner study (1999). However, there is no evidence of a correlation between size of authorship and number of publications. Therefore, since professors – who tend to publish with multiple co-authors – do not publish more on average, individual publication output has not been weighted with numbers of co-authors.

[18] Due to our limited database, methods accounting for different types of research output (see De Witte und Rogge 2010) can't be applied.

explain how many publications a person who publishes will have, need not necessarily be the same explanatory factors. To verify robustness and for estimation models in which the use of hurdle models was not suitable, (zero-inflated) negative binomial models, Poisson models and ordered probit models were estimated in addition. Rank correlations were calculated to investigate bivariate correlations between different dependent variables. The advantage of rank correlations over correlation estimations based on exact figures is that the influence of the various values is constant and large values do not have more influence. To address dependency of research performance on individuals in the same department, clusters were used for the departments in the regressions.

4. Correlations between different research performance indicators

This section looks at correlations between indicators from each of the two databases (Web of Science and Google Scholar) and correlations between the various indicators for quantitative output and outcome (citation impact).

4.1. Comparison of research performance indicators on the basis of Web of Science and Google Scholar

Publication output: number of publications

The Thomson Reuters database contains 374 publications by the 51 educational research professors. 218 of those publications (58%) are actual journal papers (articles). The remaining publications are book reviews, editorials and other forms of publication that are not relevant to our analysis. Hence, approximately 4.3 articles are included per professor on average. However, there are substantial differences in inter-individual performance. 20 percent of the professors have no articles in the database. The median figure is two articles. The mean number of articles in the Web of Science per researcher for 10 working years is two (median: one article). Analysis of publication output limited to the last six years (2005-2010) gives a mean of 1.6 articles, which is consistent with results for the overall period.

As expected, the Google Scholar database contains much more publications than Web of Science. A total of 1,559 titles are indexed, i.e., about four times the number contained in the Thomson Reuters Web of Science. The average number of publications per professor is 30.6, with a median of 22 publications. This translates as 15.8 publications per 10 working years per researcher (median: 12.7). The research output figures are slightly higher than the output identified earlier by Hornbostel and Keiner (2002).

The number of journal articles contained in both databases is 122. A surprisingly low 56 percent of Web of Science articles are also indexed in Google Scholar (conversely, 7.8 percent of Google Scholar publications are also in Web of Science). The overlap is somewhat larger if the analysis is restricted to publications since 2005 (65% and 11.8%, respectively). Nevertheless, the findings plainly show two things: firstly – unsurprisingly – the observations confirm that Web of Science contains no more than a small percentage of total publication activity (see Corby 2001, Dees 2008, Meho and Yang 2007, Moed 2005, Togia and Tsigilis 2006). Secondly, it is evident that Google Scholar contains no more than two-thirds of the "quality assured" literature given in Web of Science (see Meho and Yang

2007), which is surely an astonishingly low rate. Separate analysis of the two language regions shows marked differences between German-speaking and French-speaking institutions in terms of inclusion in Web of Science and Google Scholar. The percentage of Web of Science articles in the Google Scholar database is higher for French-speaking institutions (76%) than for German-speaking institutions (45%).

For descriptive analysis of the relationship between the number of publications per professor in the two databases, we calculated a rank correlation between the two indicators. The rank correlation is positive (r=0.46) and statistically significant at a 5% level of significance. Investigation of the relationship between Web of Science and Google Scholar using regression analysis reveals an effect size (without control variables) of approximately 0.5 percent at a 1% level of significance. This means that a 1 percent increase in publication count in Web of Science is accompanied by a 0.5 percent increase in the Google Scholar count. In absolute figures, this means that researchers have 7 publications more in Google Scholar for every 2 publications more in Web of Science.[19] This correlation does not weaken when additional control variables are included, but rises to an effect size of about 0.7 percent. The findings hence confirm that individuals who have published more journal articles that have been included in Web of Science are also more likely to have published more works elsewhere. This finding agrees with those of Puuska (2010) and Bernauer and Gilardi (2010) for political science in Switzerland, which suggest positive correlations between publication outputs in publications of different types.

Publication outcome: citation count

The 218 articles included in Web of Science altogether elicited 804 citations (not including 78 self-cites), which corresponds to an average citation frequency of 15.7 citations per professor and 3.7 citations per publication.

Like the number of publications, the citation count is also very unequally distributed. Alongside the 20 percent of professors with no published articles, another 27 percent of published professors have never produced a single article that was ever cited in an indexed publication.

The citation count in Google Scholar is also many times greater than in the Thomson Reuters database. The total citation count amounts to 12,280.[20] The average citation count is 241 citations per professor and 7.9 citations per publication. Hence, less than half of the larger citation volume in Google Scholar is attributable to the larger number of publications. The higher citation impact in Google Scholar is consistent with the results of other studies (Paludkiewicz and Wohlrabe 2010, van Aalst 2010) and can be explained by that fact that the number of possible recipients of publications contained in Google Scholar is incomparably larger than that of publications in Web of Science.

As with publication output, again we compared the correlation between the two databases as regards citations per professor. Rank correlation in this case points to an even stronger, statistically significant correlation (r=0.60) than for the publications. This strongly positive correlation is also evident in regression analysis, both with and without other control variables. Researchers with a 1 percent higher citation count per publication in Web of Science have a 0.66 percent to 0.72 percent higher citation rate per publication in Google Scholar. Therefore, as already observed in the analysis

[19] The correlation persists if the publications published in both databases are taken out of the calculations.
[20] The self-citation rate is unknown. If it is similar to the self-citation rate in Web of Science, the associated bias is negligible.

of publications, authors with high citation rates for their Web of Science publications display high citation rates for the publications in Google Scholar as well.

In summary, the conclusion is that there is a positive correlation between the Web of Science and Google Scholar indicators not only in terms of quantity but also in terms of the elicited response to research papers.

4.2. Quantity (number of publications) versus response (citation impact): comparison of the various indicators

This section looks at correlations between publication output and citation impact in educational research. Both positive and negative relationships are conceivable. A negative relationship would develop if publication quantity and quality (measured in terms of citation impact) were found to compete with each other. The relationship might be positive if good researchers were superior both in terms of output and impact, or if the probability of a researcher being noticed and cited more frequently were found to increase with the person's publication rate. The correlations of interest here are again investigated using rank correlations and multivariate estimation models.

Very high positive rank correlation coefficients (Web of Science: r=0.85, Google Scholar: r=0.85) were observed as regards the relationship between number of publications and citation count for both databases, i.e., researchers who are prolific publishers also have more citations. The strongly positive relationship is not only because the possibility of being cited increases in keeping with an increasing number of publications, since there is also a positive correlation (although less strong, especially for Google Scholar) between the number of publications and the citation count per publication (Web of Science: r=0.70, Google Scholar: r=0.45). The strong correlation between publication rate and h-index (Web of Science: r=0.88, Google Scholar: r=0.89) points in the same direction. The correlations identified here thus support the hypothesis that professors who publish more frequently also tend to author publications that elicit a higher level of response (i.e., with a higher citation impact), and refute the competition hypothesis, which posits a choice between a large number of low-impact publications and a small number of good publications with a large impact.

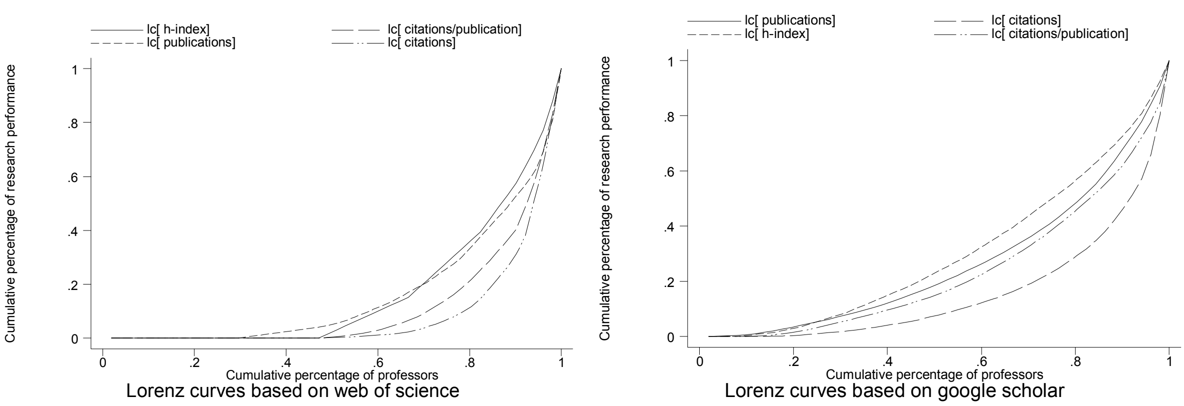
The positive relationship between a researcher's output and outcome as identified by rank correlation is confirmed in multivariate models, i.e., the positive correlation is not fuelled by individual groups of researchers with similar characteristics or through specific departments.

4.3. Distribution of research performance: major disparity between researchers

What has been said so far already gives some indication that research performance in our sample is unequally distributed between the researchers involved. To enable a more in-depth investigation of the disparities, the number of publications, citation count, citation count per publication and h-index are analysed by Lorenz curves in the following. Figure 1 clearly shows that research performance varies greatly among education science professors in Switzerland. There are a great many people with a low publication rate (0) and a handful of people who publish a great deal. A similar situation applies as regards citation count, citation count per publication, and h-index. Inequality of distribution is highest for the citation rate. The Gini index is 0.82 (Web of Science, left graph) and 0.68 (Google Scholar, right graph). By comparison, the Gini coefficients for number of publications are much lower (Web of Science: g=0.67, Google Scholar: g=0.47). The two graphs also show that the relative arrangement of the Lorenz curves is quite consistent whichever database is involved.

Comparison of the Lorenz curves for the two databases shows that, as expected, the inequality between researchers is more marked in Web of Science than Google Scholar. This finding is consistent with the fact that inclusion of publications in the Thomson Reuters is highly selective and based on rigid criteria. The inequality of distribution is compounded by the fact that 20 percent of professors do not have a single publication and 47 percent do not have a single citation in Web of Science.

Fig. 1: Lorenz curves: distribution of research performance



The highly disparate distribution of research performance inevitably raises the question as to what is causing these differences. This is investigated in the following section and an attempt is made to identify the underlying influencing factors.

5. Explanatory factors in research performance

This section presents a more in-depth investigation of the various individual and institutional factors influencing the research output and outcome of the 51 professors. The first part presents explanatory factors for Web of Science-based research performance and the second part presents the corresponding factors for Google Scholar.

5.1. Web of Science-based results

We use hurdle models to explain variance in research performance based on Web of Science data, since, given the large number of researchers who are unpublished or uncited in Web of Science, the question as to whether a person has any publications/citations and if so, how many, actually involves two separate questions, which might be influenced in different ways by the explanatory variables. Table 1 shows three models for estimation of number of articles and h-index. The three different models assessed vary in terms of influence of institutional control variables. Results are largely consistent for the two indicators investigated, i.e. number of articles and h-index.

Table 1: Hurdle regression analysis: research performance in Web of Science

	Number of articles			h-index		
	M1	M2	M3	M1	M2	M3
		hurdle (0/1)			*hurdle (0/1)*	
No. of years since PhD	-0.389	-0.161	-0.238	-0.691*	-0.057	-0.22
	(0.339)	(0.238)	(0.223)	(0.328)	(0.275)	(0.259)
No. of years since PhD, squared	0.009	0.006	0.007	0.016+	0.003	0.006
	(0.007)	(0.007)	(0.006)	(0.009)	(0.008)	(0.007)
Age	-0.947**	-1.017*	-1.184*	1.071	-0.363	0.024
	(0.357)	(0.477)	(0.497)	(1.054)	(0.949)	(0.969)
Age squared	0.008*	0.009+	0.011*	-0.011	0.003	0.000
	(0.004)	(0.005)	(0.005)	(0.010)	(0.009)	(0.009)
Female	-1.446**	-1.792**	-2.246**	-0.639	-1.294	-1.030*
	(0.470)	(0.378)	(0.713)	(0.599)	(0.884)	(0.402)
Tenured professor	4.903*	3.623**	3.767**	5.219**	2.599*	3.887*
	(2.067)	(1.353)	(1.299)	(1.821)	(1.025)	(1.806)
French language region		-3.378*			-2.933**	
		(1.365)			(1.048)	
No. of prof. colleagues (log.)		0.493			0.959	
		(0.385)			(0.673)	
Productivity of departmental colleagues			0.112			1.418*
			(0.106)			(0.679)
		neg binomial			*poisson*	
No. of years since PhD	0.302+	0.709**	0.511**	0.310*	0.533**	0.430+
	(0.167)	(0.123)	(0.139)	(0.136)	(0.151)	(0.226)
No. of years since PhD, squared	-0.005	-0.013**	-0.009**	-0.007*	-0.011**	-0.010+
	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)	(0.005)
Age	-0.494	-1.569*	-0.578	-1.413*	-1.884**	-1.639**
	(0.568)	(0.632)	(0.427)	(0.556)	(0.529)	(0.588)
Age squared	0.004	0.014*	0.005	0.013*	0.017**	0.015**
	(0.005)	(0.006)	(0.004)	(0.005)	(0.005)	(0.005)
Female	-0.721	-0.72	-1.156+	-0.691*	-1.124+	-1.462**
	(0.563)	(0.737)	(0.608)	(0.304)	(0.675)	(0.551)
Tenured professor	2.376**	1.564**	0.991+	13.533**	0.579	0.281
	(0.490)	(0.462)	(0.514)	(0.268)	(0.494)	(0.610)
French language region		-2.082**			-1.433**	
		(0.385)			(0.527)	
No. of prof colleagues (log.)		0.540			0.336	
		(0.461)			(0.333)	
Productivity of departmental colleagues			0.009			0.11
			(0.031)			(0.255)
KV: Departments	yes			yes		
KV: Areas of studies		yes	yes		Yes	yes
N	51	51	51	51	51	51

Number of articles: negative binomial-logit hurdle regression. h-index: Poisson-logit hurdle regression. Cluster for departments. Robust standard error in parentheses.
Levels of significance: + $p<0.10$, * $p<0.05$, ** $p<0.01$

The results – especially for publication output – can be summarized thus: the question as to whether a person has any publications whatsoever in Web of Science is determined to a significant extent by biological age, gender and academic position (hurdle 0/1). Younger male researchers with a tenured professorship who teach at a university in the German-speaking part of Switzerland are significantly more likely to be in Web of Science. Investigation of the other question – i.e., which factors determine how much a person who is in Web of Science actually publishes – shows that biological age gives way to years since obtaining a PhD. This confirms the recent results of Shin and Cummings (2010), who likewise identified positive effects of academic age and negative effects of biological age. The fact that biological age determines the hurdle (0/1), while "research age" determines the extent of measured research output and outcome, is an indication that publications in Web of Science may involve a cohort effect, in that older generations of researchers were under less pressure in their day to publish in foreign (indeed, English-language) journals. As far as number of publications is concerned, the positive effect of academic age should not be interpreted as a mere cumulative effect of research production. Rather, the (primarily) positive curvilinear correlation which also applies in respect of number of publications between 2005 and 2010 indicates that researchers become more efficient as they gain experience (although the efficiency benefit plateaus somewhat with time). One explanation is that researchers who have built up a rich body of knowledge and skills need less time to author a new publication. Another possibility is that individuals with more research experience have a more effective network that might work in favour of publication activity (joint publications, co-authorships).[21]

The gender effect is more of a determinant in terms of publication output in the hurdle model and in terms of citations in the negative binomial (or Poisson) part of the model. Female researchers are less likely to have publications and citations in Web of Science, and those who are represented in the database have a lower publication and citation count than their male colleagues. The poorer performance of female professors in bibliometric analyses versus their male colleagues corroborates existing findings (Aaltojärvi et al. 2008, Kyvik 1996, Puuska 2010, Rauber and Ursprung 2008) for other countries and other fields. However, gender differences as regards citation count, citation count per publication, and h-index (Web of Science) are attributable at least in part to the effect of isolated statistical outliers and lose (some) statistical significance when these are controlled for. More differentiated gender analyses that address interaction with age reveal the following findings for Web of Science data: the gender disparity varies with (academic) age. The differences are much greater for older individuals and are largely absent among younger researchers. Another indication that the disparities have declined (or disappeared altogether) in recent years is that there is no gender difference in respect of the number of publications published during the past 6 years. Hence, the gender effect is in itself a cohort effect to a very large extent.

As mentioned in the hypotheses, the higher likelihood for researchers in tenured positions to be represented in Web of Science in the first place, and to have a larger number of publications, is not amenable to a direct causal interpretation, and the data do not support a form of assessment that would result in a causal interpretation. However, the results can be taken to be robust since both biological and academic age are controlled for.

The statistically corroborated larger presence of researchers from the German-speaking part of Switzerland is attributable to a variety of factors and does not necessarily indicate that researchers from the German-speaking part of Switzerland have a higher research output and impact. German-

[21] Another possible explanation might be that long-standing researchers benefit from structural privileges due to their increasing fame / reputation (e.g. as regards allocation of research funding or inclusion of an article in a journal due to a position on the editorial board).

speaking researchers are likely to benefit from the demonstrable fact that educational research journals from German-speaking areas (specifically, Germany) are represented better in Web of Science than journals from French-speaking areas (specifically, France). German-language researchers can choose from a variety of Web of Science-indexed educational research journals: Zeitschrift für Pädagogik (since 1976), Zeitschrift für Erziehungswissenschaften (since 2006), Zeitschrift für Soziologie der Erziehung und Sozialisation (since 2005), Pädagogische Rundschau (1966-1983), Psychologie in Erziehung und Unterricht (since 1973), Zeitschrift für Entwicklungspsychologie und pädagogische Psychologie (since 1969) and Zeitschrift für pädagogische Psychologie (since 1994).[22] In contrast, important French-language publishing outlets are not represented in Web of Science. The three education sciences categories in Web of Science ("education & educational research", "special education" and "educational psychology") do not contain any French-language journals at present. Key French-language journals such as Revue française de pédagogie, Revue des sciences de l'education, and Bulletin de psychologie are not indexed.[23] The differences between the language regions may, however, be at least partly due to differences in disciplinary and communication cultures (see Keiner 1999, Späni, Hofstetter, and Schneuwly 2002).

The correlations already described also apply (or at least tend to apply) in respect of number of articles between 2005 and 2010, citation count, and citation count per article (see Appendix, Tables 6-8). Citation effects (in particular, citation count per publication) however tend to be non-statistically significant or to be lacking in robustness. In addition to possible theoretical explanations, the most likely reason has to do with methodology: as almost half of the professors have no citations to their name, variance is low and the probability of identifying significant differences commensurately remote.

The control variable results suggest that research performance may vary significantly between departments. Some disparity between individual areas of studies is also evident. Department size and the research performance of colleagues in the department have no significant influence in most instances; however, positive correlations are evident in isolated cases.

5.2. Google Scholar-based results

Table 2 presents regression results (OLS) for the Google Scholar-based number of publications and h-index. The first thing to notice is that there are fewer (if any) significant explanations for disparity in research performance to be found here than in the analyses on the basis of Web of Science data. One explanation is that the low-threshold inclusion of titles in the Google Scholar promotes statistical noise, which eliminates the statistical correlations between dependent and independent variables.[24]

[22] Moreover, there are a number of other German-speaking journals with a sociological, psychological or politological bent which also publish articles on education issues.

[23] As with the German language, a number of French-language journals specializing in related disciplines also publish articles with an education theme.

[24] Van Aalst (2010)'s findings indicate however that the obscured correlations (due to background noise) may be partly reduced by information about the specific types of publication (books, book chapters, dissertations, conference papers). This paper does not provide a more detailed attribution of Google Scholar publications because, firstly, attribution to a specific type of publication in itself tends to be the consequence of an arbitrary decision, and, secondly, not all of the links in Google Scholar actually enable access to a document (which however would be necessary for attribution to a specific form of publication).

On the other hand, (artificial) correlations promoted by the selective inclusion of journals in Web of Science should play less of a role in the Google Scholar analyses.

Estimation of the number of publications between 2005 and 2010, number of publications per author, citation count and citation count per publication reveals very similar effects (see Appendix, Tables 6-8). However, significant effects are barely in evidence for citation count per publication and number of publications between 2005 and 2010. The only exception for the latter is the academic position, and for the former it is gender: female professors achieve fewer cites per publication than their male colleagues.

Table 2: OLS regressions: research performance in Google Scholar

	Number of publications			h-index		
	M1	M2 *OLS*	M3	M1	M2 *OLS*	M3
No. of years since PhD	0.046*	0.027	0.027	0.027+	0.019	0.019
	(0.019)	(0.020)	(0.023)	(0.012)	(0.012)	(0.011)
Age	0.253	0.294	0.273	0.226*	0.279+	0.276
	(0.197)	(0.264)	(0.286)	(0.096)	(0.137)	(0.164)
Age squared	-0.003	-0.003	-0.003	-0.002*	-0.003+	-0.003
	(0.002)	(0.002)	(0.003)	(0.001)	(0.001)	(0.002)
Female	-0.039	-0.008	0.122	0.086	0.074	0.171
	(0.163)	(0.197)	(0.245)	(0.190)	(0.168)	(0.201)
Tenured professor	0.495	0.765+	0.742	0.435	0.489+	0.475+
	(0.477)	(0.385)	(0.401)	(0.320)	(0.227)	(0.217)
French language region		0.924+			0.465	
		(0.403)			(0.380)	
Number of prof colleagues (log.)		-0.094			-0.043	
		(0.188)			(0.165)	
Productivity of departmental colleagues			0.016*			0.013
			(0.006)			(0.023)
KV: Departments	Yes			yes		
KV: Areas of studies		yes	yes		yes	Yes
Adj. R-squared	0.397	0.243	0.154	0.210	0.149	0.096
N	51	51	51	51	51	51

OLS regression (DV log.). Cluster for departments. Robust standard error in parentheses.
Levels of significance: + p<0.10, * p<0.05, ** p<0.01

Analysis of the control variables detects significant disparity between departments. Differences between areas of studies are also evident in some cases. The number of colleagues in the department has no effect on individual research performance. Productivity of departmental colleagues is associated with positive point estimates in isolated cases.

6. Conclusions

Our study of the research output of educational research professors in Switzerland reveals four main results that are significant with regard to the further use of bibliographic information to assess the research performance of scholars:

Firstly, positive correlations are evident across all indicators of research performance from different bibliometric databases (Web of Science and Google Scholar). It is possible to conclude at the very least that scholars with good research performance results based on one database will also tend to do well in a measurement based on another bibliometric database, even if items counted in both databases are excluded.

Secondly, whatever the bibliometric database employed, there are positive correlations between output (number of publications) and outcome (citations), or quality and quantity. This means that the occasionally posited trade-off between quantity and quality does not apply. On the contrary: a person with a lot of publications to his or her name generally also achieves a higher impact rating with his or her publications (this also applies to the citation count per published publication).

Thirdly, explanatory models for variance in research performance, which are consistent and compatible with the existing literature, are evident in respect of Web of Science. However, the same analyses on the Google Scholar database identify virtually no statistically significant explanatory factors, indicating that the very low-threshold inclusion of publications and citations in Google Scholar makes it impossible to find explanations for the great disparities in individual publication rates that are also evident in Google Scholar.

Fourthly, there is evidence to corroborate the view that some of the explanatory models used to explain differences in research output in Web of Science may be attributable to factors that lead to unjustified researcher rankings. The main factor is certainly the language bias in the inclusion of journals in the Social Sciences Citation Index, which was found in this paper to be to the significant detriment of researchers from the French-speaking part of Switzerland.[25]

In summary, the conclusion is that evaluation of the research performance of educational research scholars on the basis of bibliometric data is justified provided that the bibliometrics are not too indiscriminate in terms of the quality of the material included. On the other hand, it needs to be ensured that the qualitative exclusion criteria do not result in a publication inclusion bias that is not justified on quality grounds and in that way works to the disadvantage of specific categories of researchers.

Finally, it is worth noting that the available bibliometric information very clearly shows that professors of educational research break down into two categories, i.e., frequently published and frequently cited researchers versus researchers who publish little to nothing and are mostly uncited. It would now be interesting to establish whether there is a trade-off between individual research performance and any other activity in higher education, e.g. whether professors who are prolific publishers devote less time to teaching or expert review activities, or whether there is no such trade-off in these areas either.

[25] This conclusion would be refutable only if French-speaking scientific journals were less well represented in SSCI than German-speaking journals for qualitative reasons.

References

Aaltojärvi, I., I. Arminen, O. Auranen, and H.-M. Pasanen. 2008. Scientific Productivity, Web Visibility and Citation Patterns in Sixteen Nordic Sociology Departments. *Acta Sociologica* 51, no. 1:5–22.

Adler, R., J. Ewing, and P. Taylor. 2009. Citation Statistics. *Statistical Science* 24, no. 1:1–14.

Archambault, É., É. Vignola-Gagne, G. Côté, V. Larivière, and Y. Gingras. 2006. Benchmarking Scientific Output in the Social Sciences and Humanities: The Limits of Existing Databases. *Scientometrics* 68, no. 3:329–42.

Bernauer, T., and F. Gilardi. 2010. Publication Output of Swiss Political Science Departments. *Swiss Political Science Review* 16, no. 2:279–303.

Bonaccorsi, A., and C. Daraio. 2003. Age Effects in Scientific Productivity: The Case of the Italian National Research Council (CNR). *Scientometrics* 58, no. 1:49–90.

Bornmann, L., R. Mutz, and H.-D. Daniel. 2008. Are There Better Indices for Evaluation Purposes than the h Index? A Comparison of Nine Different Variants of the h Index Using Data from Biomedicine. *Journal of the American Society for Information Science and Technology* 59, no. 5:830–37.

Botte, A. 2007. Scientometric Approaches to Better Visibility of European Educational Research Publications: A State-of-the-art-report. *European Educational Research Journal* 6, no. 3:303–11.

Budd, J. M. 1988. A Bibliometric Analysis of Higher Education Literature. *Research in Higher Education* 28, no. 2:180–90.

Budd, J. 1990. Higher Education Literature: Characteristics of Citation Patterns. *Journal of Higher Education* 61, no. 1:84–97.

Budd, J. M., and L. Magnuson. 2010. Higher Education Literature Revisited: Citation Patterns Examined. *Research in Higher Education* 51, no. 3:294–304.

Carayol, N., and M. Matt. 2006. Individual and Collective Determinants of Academic Scientists' Productivity. *Information Economics and Policy* 18, no. 1:55–72.

Corby, K. 2001. Method or Madness? Educational Research and Citation Prestige. *Libraries and the Academy* 1, no. 3:279–88.

Corby, K. 2003. Constructing Core Journal Lists: Mixing Science and Alchemy. *Libraries and the Academy* 3, no. 2:207–17.

Cusin, C., S. Grossenbacher, and U. Vögeli-Mantovani. 2000. *FER-Studie "Prospective de la recherche en éducation en Suisse": Teilstudie Erziehungswissenschaften an Schweizer Universitäten (Orientierung, Produktivität und Nachwuchsförderung). SKBF:* Aarau.

Davarpanah, M. R. 2009. The International Publication Productivity of Malaysia in Social Sciences: Developing a Scientific Power Index. *Journal of Scholarly Publishing* 41, no. 1:67–91.

De Witte, K., and N. Rogge. 2010. To Publish or Not to Publish? On the Aggregation and Drivers of Research Performance. *Scientometrics* 85, no. 3:657–80.

Dees, W. 2008. Innovative Scientometric Methods for a Continuous Monitoring of Research Activities in Educational Science: In: Fourth International Conference on Webometrics, Informetrics and Scientometrics & Ninth COLLNET Meeting, Berlin, 28 July -1 August 2008:1–10.

Earp, V. J. 2010. A Bibliometric Snapshot of The Journal of Higher Education and Its Impact on the Field. *Behavioral & Social Sciences Librarian* 29, no. 4:283–95.

European Commission. 2010. *Assessing Europe's University-Based Research: Expert Group on Assessment of University-Based Research. RTD.C4. EUR 24187 EN*.

Fairbairn, H., A. Holbrook, S. Bourke, G. Preston, R. Cantwell, and J. Scevak. 2009. A Profile of Education Journals. In *AARE 2008 Conference Papers Collection*, ed. P. Jeffrey.

Fernández-Cano, A., and Á. Bueno. 1999. Synthesizing Scientometric Patterns in Spanish Educational Research. *Scientometrics* 46, no. 2:349–67.

Fröhlich, G. 1999. Das Messen des leicht Messbaren. Output-Indikatoren, Impact-Masse: Artefakte der Szientometrie? In *Kommunikation statt Markt: Zu einer alternativen Theorie der Informationsgesellschaft*, ed. Jörg Becker and Wolf Göhring, 27–38. GMD Report. Sankt Augustin.

García-Pérez, M. 2010. Accuracy and Completeness of Publication and Citation Records in the Web of Science, PsycINFO, and Google Scholar: A Case Study for the Computation of h Indices in Psychology. *Journal of the American Society for Information Science and Technology* 61, no. 10:2070–85.

Gonzalez-Brambila, C., and F. M. Veloso. 2007. The Determinants of Research Output and Impact: A Study of Mexican Researchers. *Research Policy* 36, no. 7:1035–51.

Graber, M., A. Launov, and K. Wälde. 2008. Publish or Perish? The Increasing Importance of Publications for Prospective Economics Professors in Austria, Germany and Switzerland. *German Economic Review* 9, no. 4:457–72.

Haddow, G., and P. Genoni. 2010. Citation Analysis and Peer Ranking of Australian Social Science Journals. *Scientometrics* 85, no. 2:471–87.

Hall, B. H., J. Mairesse, and L. Turner. 2005. Identifying Age, Cohort and Period Effects in Scientific Research Productivity: Discussion and Illustration Using Simulated and Actual Data on French Physicists. *NBER Working Paper Series* 11739. http://www.nber.org/papers/w11739 (accessed April 13, 2011).

Harzing, A.-W. K., and R. van der Wal. 2008. Google Scholar as a New Source for Citation Analysis. *Ethics in Science and Environmental Politics* 8:61–73.

Hicks, B. 1999. The Difficulty of Achieving full Coverage of International Social Science Literature and the Bibliometric Consequences. *Scientometrics* 44, no. 2:193–215.

Hofmeister, R., and H. W. Ursprung. 2008. Das Handelsblatt Ökonomen-Ranking 2007: Eine kritische Beurteilung. *Perspektiven der Wirtschaftspolitik* 9, no. 3:254–66.

Hornbostel, S., and E. Keiner. 2002. Evaluation der Erziehungswissenschaft. *Zeitschrift für Erziehungswissenschaft* 5, no. 4:634–53.

Jacsó, P. 2008. Google Scholar revisited. *Online Information Review* 32, no. 1:102–14.

Jokić, M., and R. Ball. 2006. *Qualität und Quantität wissenschaftlicher Veröffentlichungen: Bibliometrische Aspekte der Wissenschaftskommunikation.* Schriften des Forschungszentrums Jülich 15. Jülich.

Keiner, E. 1999. *Erziehungswissenschaft 1947-1990: Eine empirische und vergleichende Untersuchung zur kommunikativen Praxis einer Disziplin.* Weinheim: Deutscher Studien Verlag.

Klusmeyer, J., H. Reinisch, and M. Söll. 2011. Wo publizieren Berufs- und Wirtschaftspädagogen? *Zeitschrift für Berufs- und Wirtschaftspädagogik* 107, no. 3:328-50.

Knorr, K. D., R. Mittermeir, G. Aichholzer, and G. Waller. 1979. Individual Publication Productivity as a Social Position Effect in Academic and Industrial Research Units. In *The Effectiveness of Research Groups in Six Countries*, ed. Frank M. Andrews, 55–94. Cambridge: Cambridge University Press.

Krampen, G., R. Becker, U. Wahner, and L. Montada. 2007. On the Validity of Citation Counting in Science Evaluation: Content Analyses of References and Citations in Psychological Publications. *Scientometrics* 71, no. 2:191–202.

Kroc, R. J. 1984. Using Citation Analysis to Assess Scholarly Productivity. *Educational Researcher* 13, no. 6:17–22.

Kyvik, S. 1996. Child Care, Research Collaboration, and Gender Differences in Scientific Productivity. *Science, Technology & Human Values* 21, no. 1:54–71.

Larivière, V., É. Vignola-Gagné, C. Villeneuve, P. Gélinas, and Y. Gingras. 2011. Sex Differences in Research Funding, Productivity and Impact: An Analysis of Québec University Professors. *Scientometrics* 87, no. 3:483–98.

Levin, S. G., and P. E. Stephan. 1991. Research Productivity over the Life Cycle: Evidence for Academic Scientists. *American Economic Review* 81, no. 1:114–32.

Linmans, A. J. M. 2010. 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. *Scientometrics* 83, no. 2:337–54.

Luce, T. S., and D. M. Johnson. 1978. Rating of Educational and Psychological Journals. *Educational Researcher* 7, no. 10:8–10.

Meho, L. I., and K. Yang. 2007. Impact of Data Sources on Citation Counts and Rankings of LIS Faculty: Web of Science versus Scopus and Google Scholar. *Journal of the American Society for Information Science and Technology* 58, no. 13:2105–25.

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, no. 1:81–100.

Nederhof, A. J., M. Luwel, and H. F. Moed. 2001. Assessing the Quality of Scholarly Journals in Linguistics. An Alternative to Citation-based Journal Impact Factors. *Scientometrics* 51, no. 1:241–65.

Neuhaus, C. 2010. *Vergleichende Analysen von Forschungsleistungen: Forschungsgruppen im Spiegel bibliometrischer Indikatoren.* Baden-Baden: Nomos.

Paludkiewicz, K., and K. Wohlrabe. 2010. Qualitätsanalyse von Zeitschriften in den Wirtschaftswissenschaften - über Zitationsdatenbanken und Impaktfaktoren im Online-Zeitalter. *ifo Schnelldienst* 63, no. 21:18–28.

Panaretos, J., and C. Malesios. 2009. Assessing Scientific Research Performance and Impact with Single Indices. *Scientometrics* 81, no. 3:635–70.

Prathap, G. 2011. The Fractional and Harmonic p-Indices for Multiple Autorship. *Scientometrics* 86, no. 2:239–44.

Puuska, H.-M. 2010. Effects of Scholar's Gender and Professional Position on Publishing Productivity in Different Publication Types: Analysis of a Finnish University. *Scientometrics* 82, no. 2:419–37.

Rauber, M., and H. W. Ursprung. 2008. Life Cycle and Cohort Productivity in Economic Research: The Case of Germany. *German Economic Review* 9, no. 4:431–56.

Rey, O. 2009. Quality Indicators and Educational Research Publications: Which Publications Count? *Dossier d'actualité de la VST* 46. http://www.inrp.fr/vst/LettreVST/english/46-june-2009_en.php?onglet=integrale (accessed June 14, 2011).

Ritzberger, K. 2008. Eine invariante Bewertung wirtschaftswissenschaftlicher Fachzeitschriften. *Perspektiven der Wirtschaftspolitik* 9, no. 3:267–85.

Schulze, G. G., S. Warning, and C. Wiermann. 2008. Zeitschriftenrankings für die Wirtschaftswissenschaften - Konstruktion eines umfassenden Metaindexes. *Perspektiven der Wirtschaftspolitik* 9, no. 3:286–305.

Shin, E.-J. 2004. Measuring the Impact of Electronic Publishing on Citation Indicators of Education Journals. *Libri* 54, no. 4:221–27.

Shin, J. C., and W. K. Cummings. 2010. Multilevel Analysis of Academic Publishing across Disciplines: Research Preference, Collaboration, and Time on Reserach. *Scientometrics* 85, no. 2:581–94.

Silverman, R. J. 1985. Higher Education as a Maturing Field? Evidence from Referencing Practices. *Research in Higher Education* 23, no. 2:150–83.

Smart, J. C. 1983. Perceived Quality and Citation Rates of Education Journals. *Research in Higher Education* 19, no. 2:175–82.

Smart, J. C., and C. F. Elton. 1981. Structural Characteristics and Citation Rates of Education Journals. *American Educational Research Journal* 18, no. 4:399–413.

Smart, J. C., and G. W. McLaughlin. 1982. Education Specialty Areas. *Educational Researcher* 11, no. 7:10–13.

Smeby, J.-C., and S. Try. 2005. Departmental Contexts and Faculty Research Activity in Norway. *Research in Higher Education* 46, no. 6:593–619.

Späni, M., R. Hofstetter, and B. Schneuwly. 2002. Interweaving Educational Sciences and Pedagogy with Professional Education: Contrasting Configurations at Swiss Universities, 1870-1950. *European Educational Research Journal* 1, no. 1:45–64.

Tight, M. 2008. Higher Education Research as Tribe, Territory and/or Community: A Co-citation Analysis. *Higher Education* 55, no. 5:593–605.

Todeschini, R. 2011. The j-Index: A new Bibliometric Index and Multivariate Comparisons between Other Common Indices. *Scientometrics* 87, no. 3:621–39.

Togia, A., and N. Tsigilis. 2006. Impact Factor and Education Journals: A Critical Examination and Analysis. *International Journal of Educational Research* 45, no. 6:362–79.

van Aalst, J. 2010. Using Google Scholar to Estimate the Impact of Journal Articles in Education. *Educational Researcher* 39, no. 5:387–400.

van Ours, J. 2009. Will you still Need me: When I'm 64? *De Economist* 157, no. 4:441–60.

van Leeuwen, T. 2006. The Application of Bibliometric Analyses in the Evaluation of Social Science Research. Who Benefits from it, and why it is still Feasible. *Scientometrics* 66, no. 1:133–54.

Wellington, J., and C. J. Torgerson. 2005. Writing for Publication: What Counts as a 'High Status, Eminent Academic Journal'? *Journal of Further and Higher Education* 29, no. 1:35–48.

Winkelmann, R. 2008. *Economic Analysis of Count Data.* Fifth edition. Heidelberg, New York: Springer.

Appendix

Table 3: Descriptive statistics of variables: means and variances

Variable	Obs	Mean	Std. Dev.	Min	Max
Indicators based on Web of Science					
Number of articles	51	4.27	7.16	0	42
Number of articles 2005-2010	51	1.57	2.62	0	12
Citation count (not including self-cites)	51	15.84	34.73	0	151
Citation count/article	51	1.97	3.68	0	16.88
h-index	51	1.29	1.80	0	8
Indicators based on Google Scholar					
Number of publications	51	30.57	29.23	1	144
Number of publications/author	51	19.00	17.33	0.33	73.63
Number of publications 2005-2010	51	8.65	8.79	0	47
Citation count	51	240.78	437.65	0	2412
Citation count/publication	51	6.40	7.95	0	49.50
h-index	51	5.61	4.36	0	21
Individual characteristics					
Number of years since PhD	51	20.22	8.20	6	39
Age	51	54.90	7.15	39	65
Gender (female)	51	0.35	0.48	0	1
Prof. category: tenured professorship	51	0.61	0.49	0	1
Institutional/structural characteristics					
Departments					
Reference: Basel, St. Gallen, IFE Zurich	51	0.14	0.35	0	1
IGB Zurich	51	0.06	0.24	0	1
Berne	51	0.10	0.30	0	1
PEDG Fribourg	51	0.08	0.27	0	1
LB Fribourg	51	0.08	0.27	0	1
Neuchâtel	51	0.08	0.27	0	1
Geneva	51	0.47	0.50	0	1
French language region	51	0.63	0.49	0	1
No. of prof colleagues	51	12.43	10.09	0	23
Area of studies					
Didactics	51	0.29	0.46	0	1
General pedagogy	51	0.18	0.39	0	1
Adult education	51	0.14	0.35	0	1
Sociology, systems research	51	0.18	0.39	0	1
Child psychology, special needs teaching, anthrop.	51	0.22	0.42	0	1

Table 4: Descriptive account of variables: quartiles and totals

	total	mean	p25	p50	p75
Number of articles (WoS)	218	4.3	0	2	5
Number of articles 2005-2010 (WoS)	80	1.6	0	0	2
Number of publications (GS)	1559	30.6	12	22	41
Number of publications 2005-2010 (GS)	441	8.6	2	7	12
Number of articles in both databases	122	2.4	0	1	3
Number of articles 2005-10	52	1.0	0	0	1
Percentage with articles (GS)	7.8	8.7	0	4.0	12.3
Percentage with articles (WoS)	56.0				
Percentage with articles in 05-10 (GS)	11.8	13.1	0	0	14.3
Percentage with articles in 05-10 (WoS)	65.0				
Citation count (WoS)	808	15.8	0	1	12
Citation count (GS)	12280	240.8	35	99	253
Citation count/article (WoS)		2.0	0	0.3	2.4
Citation count/publication (GS)		6.4	2.5	3.7	8.7

Table 5: Descriptive statistics of variables: Gini index

	Gini index
Indicators based on Web of Science	
Number of articles	0.67
Number of articles 2005-2010	0.74
Citation count	0.82
Citation count/article	0.76
h-index	0.66
Indicators based on Google Scholar	
Number of publications	0.47
Number of publications 2005-2010	0.50
Citation count	0.68
Citation count/publication	0.53
h-index	0.40

Table 6: Regression results: number of publications 2005-2010

	Web of Science			Google Scholar		
	M1	M2	M3	M1	M2	M3
		hurdle (0/1)			*OLS*	
No. of years since PhD	-0.177	0.356	0.042	0.006	-0.007	-0.004
	(0.371)	(0.386)	(0.244)	(0.022)	(0.021)	(0.016)
No. of years since PhD, squared	0.003	-0.008	-0.001			
	(0.008)	(0.008)	(0.005)			
Age	-0.918	-1.282	-0.408	0.01	-0.04	0.014
	(0.839)	(1.108)	(0.821)	(0.274)	(0.301)	(0.295)
Age squared	0.007	0.01	0.003	0.000	0.000	0.000
	(0.008)	(0.011)	(0.008)	(0.003)	(0.003)	(0.003)
Female	0.793	0.298	-0.101	0.146	0.312	0.335
	(0.637)	(0.543)	(0.524)	(0.284)	(0.332)	(0.335)
Tenured professor	4.975**	3.298**	2.622+	0.859	0.954+	0.913*
	(1.405)	(1.192)	(1.565)	(0.577)	(0.418)	(0.374)
French language region		-6.034**			0.306	
		(1.382)			(0.183)	
Number of prof colleagues (log.)		1.400*			-0.19	
		(0.629)			(0.122)	
Productivity of departmental colleagues			0.364			-0.027
			(0.475)			(0.019)
		poisson				
No. of years since PhD	0.411*	0.848*	0.542			
	(0.176)	(0.342)	(0.356)			
No. of years since PhD, squared	-0.013**	-0.020**	-0.013			
	(0.005)	(0.008)	(0.008)			
Age	-1.674**	-1.956**	-1.548+			
	(0.467)	(0.722)	(0.892)			
Age squared	0.016**	0.018**	0.015+			
	(0.005)	(0.007)	(0.008)			
Female	-0.364**	-0.005	-0.816**			
	(0.131)	(0.372)	(0.244)			
Tenured professor	13.887**	0.623	0.127			
	(0.135)	(0.469)	(0.388)			
French language region		-1.157**				
		(0.352)				
Number of prof colleagues (log.)		-0.026				
		(0.281)				
Productivity of departmental colleagues			-0.215			
			(0.138)			
KV: Departments	yes			yes		
KV: Areas of studies		yes	yes		yes	yes
Adj. R-squared				0.090	0.132	0.149
N	51	51	51	51	51	51

Web of Science: Poisson-logit hurdle regression. Google Scholar: OLS regression (DV log.).
Cluster for departments. Robust standard error in parentheses. + p<0.10, * p<0.05, ** p<0.01

Table 7: Regression results: citation count

	Web of Science			Google Scholar		
	M1	M2	M3	M1	M2	M3
		hurdle (0/1)			*Ordered probit*	
No. of years since PhD	-0.692*	-0.057	-0.147	0.040+	0.038*	0.034+
	(0.328)	(0.275)	(0.225)	(0.024)	(0.018)	(0.018)
No. of years since PhD, squared	0.016+	0.003	0.005			
	(0.009)	(0.008)	(0.007)			
Age	1.073	-0.363	-0.215	0.157	0.215	0.205
	(1.054)	(0.949)	(0.877)	(0.242)	(0.249)	(0.283)
Age squared	-0.011	0.003	0.002	-0.002	-0.003	-0.003
	(0.010)	(0.009)	(0.008)	(0.002)	(0.003)	(0.003)
Female	-0.639	-1.294	-1.240**	-1.454**	-1.385**	-1.068**
	(0.599)	(0.884)	(0.459)	(0.305)	(0.322)	(0.390)
Tenured professor	5.220**	2.599*	2.956*	1.476**	1.086**	0.998**
	(1.822)	(1.025)	(1.443)	(0.430)	(0.374)	(0.371)
French language region		-2.933**			1.536+	
		(1.048)			(0.908)	
Number of prof colleagues (log.)		0.959			-0.389	
		(0.673)			(0.341)	
Productivity of departmental colleagues			0.067+			0.001*
			(0.037)			(0.000)
		negative binomial				
No. of years since PhD	-0.329	0.845	0.592			
	(0.210)	(0.541)	(1.174)			
No. of years since PhD, squared	0.001	-0.013+	-0.012			
	(0.004)	(0.007)	(0.016)			
Age	-1.929*	-1.307*	-1.847**			
	(0.772)	(0.550)	(0.554)			
Age squared	0.020*	0.01	0.016**			
	(0.008)	(0.006)	(0.006)			
Female	-5.841**	-0.718	-3.37			
	(1.480)	(3.068)	(4.237)			
Tenured professor	17.204**	0.984	1.097			
	(0.686)	(1.338)	(1.762)			
French language region		-0.715				
		(0.619)				
Number of prof. colleagues (log.)		-0.939*				
		(0.426)				
Productivity of departmental colleagues			0.022			
			(0.033)			
KV: Departments	yes			yes		
KV: Areas of studies		yes	yes		yes	yes
Pseudo R-squared				0.306	0.279	0.214
N	51	51	51	51	51	51

Web of Science: Negative Binomial-Logit Hurdle Regression. Google Scholar: Ordered Probit Regression (3 cat.). Cluster for departments. Robust standard error in parentheses. + p<0.10, * p<0.05, ** p<0.01

Table 8: Regression results: citation count per publication

	Web of Science			Google Scholar		
	M1	M2	M3	M1	M2	M3
		Negative binomial			*Ordered probit*	
No. of years since PhD	-0.075	0.249	0.238+	-0.004	-0.008	-0.005
	(0.265)	(0.180)	(0.123)	(0.019)	(0.017)	(0.014)
No. of years since PhD, squared	0.003	-0.004	-0.004			
	(0.007)	(0.005)	(0.004)			
Age	-0.351	-0.692	-0.646	-0.48	-0.415	-0.469
	(1.158)	(0.775)	(0.692)	(0.382)	(0.370)	(0.329)
Age squared	0.003	0.006	0.005	0.004	0.004	0.004
	(0.011)	(0.007)	(0.006)	(0.003)	(0.003)	(0.003)
Female	-1.419**	-1.879**	-2.169**	-0.840**	-1.067*	-1.005*
	(0.336)	(0.663)	(0.503)	(0.269)	(0.477)	(0.448)
Tenured professor	3.312**	0.987	1.022	0.121	0.01	-0.174
	(0.636)	(0.659)	(0.765)	(0.277)	(0.216)	(0.311)
French language region		-0.521			0.39	
		(0.626)			(0.557)	
Number of prof colleagues (log.)		-0.247			0.086	
		(0.634)			(0.216)	
Productivity of departmental colleagues			0.109			-0.036*
			(0.129)			(0.014)
KV: Departments	yes			yes		
KV: Areas of studies		yes	yes		yes	yes
Pseudo R-squared				0.096	0.117	0.108
N	51	51	51	51	51	51

Web of Science: Negative Binomial Regression. Google Scholar: Ordered Probit Regression (3 cat.).
Cluster for departments. Robust standard error in parentheses. + p<0.10, * p<0.05, ** p<0.01