

Improving the Coverage of Social Science and Humanities Researchers' Output: The Case of the Érudit Journal Platform

Vincent Larivière

École de bibliothéconomie et des sciences de l'information, Université de Montréal, C.P. 6128, Succ. Centre-Ville, Montréal, QC. H3C 3J7, Canada; Cyberinfrastructure for Network Science Center, School of Library and Information Science, Indiana University, 10th St. & Jordan Ave., Wells Library, Bloomington, IN 47405, USA; Observatoire des sciences et des technologies (OST), Centre interuniversitaire de recherche sur la science et la technologie (CIRST), Université du Québec à Montréal. E-mail: vincent.lariviere@umontreal.ca

Benoit Macaluso

Observatoire des sciences et des technologies (OST), Centre interuniversitaire de recherche sur la science et la technologie (CIRST), Université du Québec à Montréal, CP 8888, Succ. Centre-Ville, Montréal, QC. H3C 3P8, Canada. E-mail: macaluso.benoit@uqam.ca

In non-English-speaking countries the measurement of research output in the social sciences and humanities (SSH) using standard bibliographic databases suffers from a major drawback: the underrepresentation of articles published in local, non-English, journals. Using papers indexed (1) in a local database of periodicals (Érudit) and (2) in the Web of Science, assigned to the population of university professors in the province of Québec, this paper quantifies, for individual researchers and departments, the importance of papers published in local journals. It also analyzes differences across disciplines and between French-speaking and English-speaking universities. The results show that, while the addition of papers published in local journals to bibliometric measures has little effect when all disciplines are considered and for anglophone universities, it increases the output of researchers from francophone universities in the social sciences and humanities by almost a third. It also shows that there is very little relation, at the level of individual researchers or departments, between the output indexed in the Web of Science and the output retrieved from the Érudit database; a clear demonstration that the Web of Science cannot be used as a proxy for the “overall” production of SSH researchers in Québec. The paper concludes with a discussion on these disciplinary and language differences, as well as on their implications for rankings of universities.

Introduction

The measurement of research output in the social sciences and humanities (SSH)—using standard bibliographic databases such as Thomson Reuters' Web of Science (WoS) or Elsevier's Scopus—suffers from two major drawbacks. The first is that they only index journal articles and conference proceedings and, hence, exclude books and book chapters, which account for a significant proportion of the research output in these disciplines (Hicks, 1999, 2004; Huang & Chang, 2008; Larivière, Archambault, Gingras, & Vignola-Gagné, 2006; White, Boell, Yu, Davis, Wilson & Cole, 2009). This limitation affects the various SSH disciplines in different ways: more quantitative disciplines such as economics and psychology increasingly rely on journal articles—in a manner similar to disciplines of the natural and medical sciences (NMS)—while other disciplines of the humanities, such as history and literature, continue to rely mainly on books and do not increase their use of serials (Larivière et al., 2006).

The second major drawback is that these databases overrepresent English-language literature (Archambault, Vignola-Gagné, Côté, Larivière, & Gingras, 2006). In other words, a greater proportion of the existing English-language literature is indexed in those databases, both NMS and in SSH, compared with French, German, or Chinese literature. This limitation affects non-English speaking countries directly: the lower rate of coverage of their local non-English SSH journals—which are important publication venues for research results in these disciplines—significantly reduces the proportion of their output that is included in international

Received June 16, 2011; revised July 14, 2011; accepted July 15, 2011

© 2011 ASIS&T • Published online 19 September 2011 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/asi.21632

rankings and comparisons. It is a well-known fact that research in the SSH is more local than that of the NMS—which is, by essence, international (Nederhof, 2006). Indeed, an electron behaves in the same manner in Canada, South Africa, or China, while the same cannot be said of the social structure or history of such societies. This is reflected in the publication venues of researchers: a historian of Québec will generally publish his/her findings in a French-language history journal, and these journals are not generally indexed by the standard citation indexes. On the other hand, the physicist or biochemist of the same region, whose frame of reference is an international scientific community that studies objects and phenomena that are international by essence, will much more likely publish in international journals. As a consequence, SSH researchers' production not only suffers from a lower rate of coverage compared to that of the NMS because of the nature of its main publication types, but also, in non-English-speaking countries, because of the weaker coverage of local, non-English, publication venues. The scientific production of SSH researchers from non-English-speaking countries is, thus, underestimated compared with that of their SSH colleagues from English-speaking countries, as the local SSH journals of the latter are indexed (Archambault et al., 2006). This leads to absurd international rankings of research in SSH (e.g., Godin, 2002), where the U.S. and the U.K. are the two most productive countries in SSH, followed by Canada and Australia, and then by Germany, the Netherlands, and France. For nobody can seriously argue that Canadian or Australian researchers are more active in SSH research than their German or French colleagues.

In order to increase the coverage of their research outputs, several countries—especially in the developing world—have created databases aimed at covering their “local” output. For some countries where NMS research is also badly covered, these databases are not limited to the SSH. One example of such an initiative is the Scientific Electronic Library Online (SciELO) (<http://www.scielo.org>), which was created in 1997 in collaboration between the Latin American and Caribbean Center on Health Science Information (BIREME; PAHO and WHO) and Sao Paulo State Foundation for Support to Science (FAPESP) (Meneghini, Mugnaini, & Packer, 2006). Given that this database serves as an access point for papers, it can also combine standard citation metrics with usage metrics. As of June 2011, SciELO indexed papers from 847—mainly—Latin American journals in both NMS and SSH. Researchers from other countries have also created their own databases of periodicals. Chinese periodicals are covered in the Chinese Science Citation Database (CSCD)—now included in Thomson Reuters' WoS—the Chinese Social Sciences Citation Index (CSSCI), the Chinese Humanities and Social Science Citation Database (CHSSCD), and the Chinese Scientific and Technical Papers and Citations (CSTPC) (Jin & Wang, 1999; Jin, Jiangong, Chen, & Zhu, 2002; Liang, 2003; Zhou, Su, & Leydesdorff, 2010). Similar databases were also created in Japan (Negishi, Sun, & Shigi, 2004) and South Africa (Tijssen, Mouton, van Leeuwen, & Boshoff, 2006), among others. Although analyses of the citation characteristics of

the serials contained in these databases has been performed, no study has yet used these national databases to measure how they improve the coverage of publication records at the individual researcher and departmental levels.

This paper analyzes how the addition of local journals improves the measurement of research outputs in the SSH and NMS, using papers from the WoS and from the Érudit database—a digital publishing platform containing Québec scholarly journals—assigned to the entire population of university professors in the French-speaking Canadian province of Québec. Answers to two general questions are sought. First, what is the proportion of Québec researchers' papers that is added when local journals are included? How does this vary across the spectrum of disciplines of the SSH and between French-speaking and English-speaking universities? Second, how does, at the individual researcher level, the production indexed in the WoS compare with the production published in local journals? In other words, is there a relation between the number of papers a researcher has in the WoS and the number of papers she/he publishes in local journals? The following section describes the data sources and methods. The results are then presented and discussed in the Conclusion.

Methods and Data Sources

This paper merges three sources of data. The backbone of this paper is the list of all Québec researchers (14,500 individuals), which was obtained through an agreement with Québec's *Ministère de la recherche, de l'innovation de et l'exportation* and the province's three research councils (FQRNT, FQRSC, FRSQ). This list was essential in order to assign papers to individual researchers. In addition to the full names of professors, this list also includes their universities and departments. Departments were categorized into 43 disciplinary categories based on the U.S. Classification of instructional programs (CIP).¹ The list of researchers was matched, using the surname and first initials of professors/authors with all 2000–2009 papers with at least one Québec address indexed in Thomson Reuters' WoS, as well as with 2000–2009 papers of the Érudit database. This automatic match yielded 182,463 articles and 533,599 author–article combinations, which were reduced respectively to 88,168 and 139,858 once the manual removal of papers authored by homonyms was completed.

Unlike the WoS, the Érudit (<http://www.erudit.org/?lang=en>) database is not a citation index but a web journal platform through which researchers have access to scholarly papers. It indexes, in XML format, a paper's content and metadata, including cited references—which, unfortunately, are not yet in a format that allows their use for citation analyses. The Érudit database used in this paper included 83 Québec scholarly periodicals, for a total of almost 62,000 documents, of which 11,000 are categorized as articles,

¹More details can be found in Larivière (2010) and Larivière, Macaluso, Archambault, & Gingras, (2010).

TABLE 1. Percentage of cited papers and average number of citations received in the Web of Science (2000–2010) for papers indexed in the Web of Science and in *Érudit* (2000–2005).

Domain	Web of Science		Web of Science (French Only)		Érudit	
	% Cited papers	Avg. no. citations	% Cited papers	Avg. no. citations	% Cited papers	Avg. no. citations
AH	20.0%	0.38	5.8%	0.07	8.7%	0.16
NMS	84.4%	2.96	51.2%	1.55	40.4%	1.18
SS	76.8%	3.34	29.8%	0.74	13.7%	0.28

research notes, or review articles. Thirty-four of these journals are currently funded by the FRQSC *programme de soutien aux revues de recherche et de transfert de connaissances* (<http://www.fqsc.gouv.qc.ca/fr/section.php?id=123>). The metadata of each paper indexed in *Érudit* was transferred to OST, which transformed the whole platform into a relational SQL database for bibliometric analyses. The same was done for the WoS data. In both databases only articles, notes, and review articles were retained for this analysis. In order to see the difference between the 15 French-speaking² and three English-speaking³ universities in the province, each professor was coded as belonging to either one of the two categories of universities.

Finally, in order to obtain contextual data on the scientific impact of *Érudit*-indexed journals, we have matched, for 2000–2005 papers, all of the citations they received in the WoS over the 2000–2010 period (Table 1). Similar numbers were also compiled for journals indexed in the WoS (all languages and French-language papers only). In order to be comparable, WoS data excludes journal self-citations. The table shows that, although journals indexed in WoS obtain more citations than *Érudit* journals, the difference is much smaller when only French-language papers are considered. Moreover, WoS-indexed papers in French published in arts and humanities journals obtain fewer citations than papers published in arts and humanities journals indexed by *Érudit*. This confirms the well-known fact that Thomson's indexing policy in the humanities is more subjective than in the social sciences or natural sciences; i.e., it is not solely based on citations (Archambault et al., 2006).

Results

Papers assigned to Québec researchers were categorized into three categories: (1) papers exclusively indexed in the WoS, (2) papers exclusively indexed in *Érudit*, and (3) papers indexed in both WoS and *Érudit*. On the whole, 85,185

²Université Laval, Université de Montréal, École des hautes études commerciales de Montréal, École polytechnique de Montréal, Université de Sherbrooke, Université du Québec, École nationale d'administration publique, École de technologie supérieure, Institut national de la recherche scientifique, Université du Québec en Abitibi-Témiscamingue, Université du Québec à Chicoutimi, Université du Québec en Outaouais, Université du Québec à Montréal, Université du Québec à Rimouski, and Université du Québec à Trois-Rivières.

³Bishop's University, Concordia University, and McGill University.

papers were from the WoS only, 2,926 from *Érudit* only, and 59 from both databases were assigned to 10,238 researchers, which means that 4,262 researchers have not published a single paper in either the WoS- or *Érudit*-indexed journals.⁴ There is, thus, very little overlap between papers indexed by the *Érudit* platform and by the WoS. This also shows that, at the level of all disciplines taken together, the addition of the *Érudit* database has a modest effect—a 3.4% increase—on the number of papers retrieved for Québec researchers. When only professors from French-speaking universities are considered, this increase is 5.1%, while the percentage is only 1% when only English-speaking universities are considered. This increase is more important when only SSH researchers are considered: 29.8% for those affiliated with francophone universities and of 4.8% for those affiliated with anglophone universities.

This proportion differs significantly across the spectrum of academic departments. Figure 1 presents, for departments of all universities and departments of French- and English-speaking universities, the distribution of the proportion of their papers that are in the *Érudit* database (A), as well as the number of papers in each database for departments in which more than 10% of their papers were published in *Érudit* (B). In two-thirds of the departments (29 departments), less than 10% of researchers' papers are in *Érudit*, and, for more than half (55%) of the departments (24 departments), this percentage is less than 5%. With the exception of psychology, these departments all belong to disciplines of the NMS. On the other hand, for 15 departments (34%), 10% or more of researchers' papers are in the *Érudit* database (Figure 1B). Not surprisingly, these departments are all in the SSH. As one would expect, the percentage of papers in the *Érudit* database is always greater when only French-speaking universities are considered: for five departments, more than half of the scientific production covered by this study comes from the *Érudit* database.

It is worth noting that for seven departments—Religious Studies & Vocations, Social Work, Education, French/English, Anthropology, Archaeology & Sociology, Fine & Performing Arts and Other Social Sciences & Humanities—more than 33% of the output consist of *Érudit*-indexed papers (Figure 1B). Journal articles authored by

⁴Given that only 59 papers were in both databases, data presented in the remaining of the paper are divided as WoS-indexed papers and *Érudit*-indexed papers not indexed in the WoS.

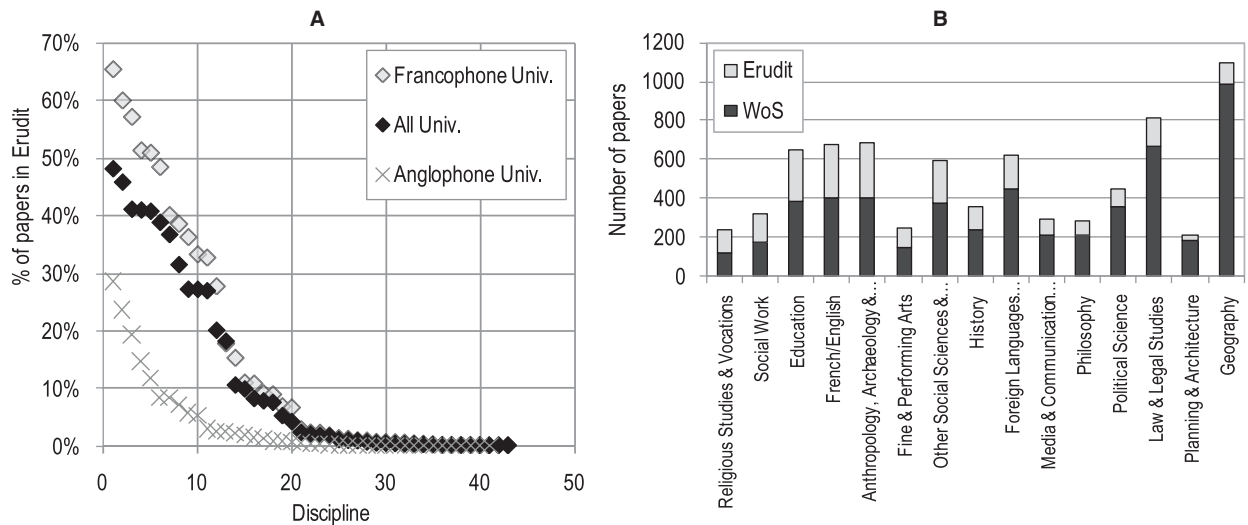


FIG. 1. (A) Distribution of the percentage of papers in the Érudit database, by departments' discipline (in decreasing order) and type of university. (B) Departments' disciplines in which more than 10% of researchers' papers are in Érudit database, 2000–2009.

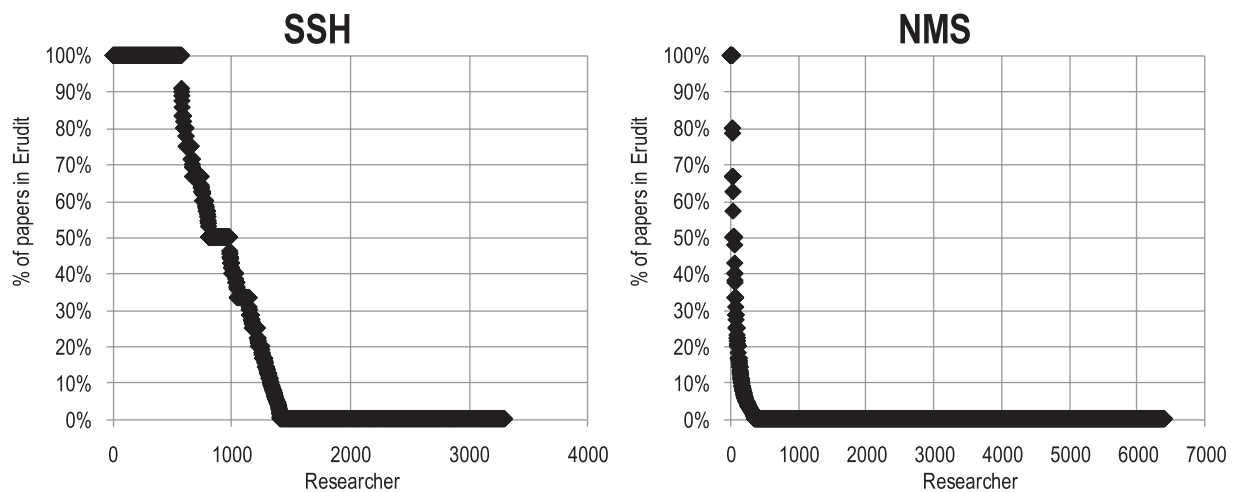


FIG. 2. Percentage of papers in the Érudit database for researchers (all universities) of the social sciences and humanities (SSH) and of the natural and medical sciences (NMS), 2000–2009.

researchers from these disciplines are thus significantly underestimated when measured exclusively by the WoS database. It is, again, even more the case when only French-speaking universities are considered.

We can distinguish three groups when data are compiled at the level of individual researchers (Figure 2). When SSH researchers from all universities are considered, 18% of them publish exclusively in Érudit-indexed journals, 25% in both Érudit and the WoS, and 57% in the WoS only. The Érudit percentages are much higher when only professors from French-language universities are analyzed: 24% of professors publish in Érudit only, 32% in both Érudit and WoS, and 44% in WoS only. In other words, more than half of the SSH professors in French-language universities have published at least one paper in an Érudit-indexed journal. For researchers of English-speaking universities, it is the opposite: 3% of professors have published at least one

paper in Érudit journals only, 11% in journals indexed in both Érudit and WoS, and 86% in WoS journals exclusively. Unsurprisingly, the importance of Érudit is much smaller for researchers of NMS (all universities): 1% publish exclusively in Érudit-indexed journals, 5% in those indexed by both Érudit and WoS, and 94% in WoS-indexed journals only. Only a slight increase is observed in NMS when only professors from French-speaking universities are analyzed.

Similarly, although SSH researchers from French-speaking universities account for two-thirds of the population studied here, they represent 94% and 87% of those who exclusively publish papers in Érudit journals and who publish in both Érudit and WoS journals, respectively. A similar tendency is also observed in NMS, with percentages of 97% and 91%, respectively. On the other hand, the one-third of SSH researchers from anglophone universities account for 6% of researchers who publish exclusively in Érudit and

13% of those who publish in both *Érudit*- and WoS-indexed journals.

Given these results, it does not come as a surprise that there is very little correlation between researchers' output profile in WoS and their research output profile in *Érudit*. Analyses at the level of all SSH or NMS researchers, be it for French- or for English-speaking universities, have not yielded any correlation. On the other hand, we have found a few correlations at the level of specific disciplines (see Appendix). In Fine & Performing Arts and Other Health Sciences, negative correlation coefficients of -0.33 and -0.51 , respectively, have been found, while positive correlations coefficients were found in French/English (0.82), Religious Studies & Vocations (0.71), and Foreign Languages Literature, Linguistics & Area Studies (0.69) and, to a lesser extent, in Earth and Ocean Sciences (0.53). Given these few "positive" correlations, one cannot use WoS-indexed papers as a surrogate for the "overall" production of SSH researchers in Québec—each database measures the output of a distinct community working on different topics.

Discussion and Conclusion

Data presented in the preceding section have shown that the addition of papers indexed in the *Érudit* platform to those indexed in the WoS results in an increase of 3.4% in the number of papers published by Québec researchers when all disciplines and universities are considered. On the other hand, the increase is more important when only SSH researchers are considered: it is 29.8% for those affiliated with francophone universities and 4.8% for those affiliated with anglophone universities. These results do not come as a surprise, as *Érudit* almost exclusively indexes journals from the SSH disciplines. As previously mentioned in this paper, there are not a lot of "local" NMS journals, nor are there many "local" SSH journals in English. There is, thus, a very important proportion—almost a third—of the article production of SSH researchers that is missed when data are limited to the WoS.

Even adding *Érudit*-indexed papers does not provide a full count of SSH researchers' papers not indexed in WoS. The *Érudit* platform covers only Québec journals and, hence, excludes (1) French journals, with which the Québec francophone research community has very important ties, as well as (2) Canadian journals outside Québec—often bilingual but with a predominance of English—in which Québec researchers from anglophone universities are more likely to publish. Although our results suggest that the anglophone community does not publish a great deal in "local" venues, they do not take into account the portion of their papers in Canadian journals from outside Québec, which are more likely to be in English. That being said, our results either mean that SSH researchers from anglophone universities (1) do not work much on local topics or (2) manage to publish their work on local topics in international journals or in Canadian journals from outside Québec.

Along the same lines, the few correlations found at the level of individual researchers and of departments between the numbers of papers indexed in the *Érudit* platform and in the WoS suggest that it is not the same group of researchers who work on local topics and those who work on topics that are more international in scope. And these groups are quite homogeneous: those who publish in international journals do not publish much in local journals, and vice versa. There is also a strong disciplinary component to these differences. As shown in Figure 1B, most of the disciplines with a greater proportion of papers indexed by *Érudit* are disciplines that have close ties to the social context (Québec) in which they function, both in terms of its institutions and political system, and also its culture (religious studies, social work, education, French/English, anthropology, archeology & sociology, fine & performing arts, etc). Contrary to the research published in international journals—which is often said to be the most visible to the international scientific community—papers published in local journals may be in many cases the part of researchers' output that is the most visible to the local community of researchers working on local topics as well as to those working on these topics outside of academia. For example, papers debating the strengths and weaknesses of the education system in Québec will be necessarily published in local journals, which are read by the actors of the education system and to which they are more likely to contribute. Researchers in more quantitative domains, such as psychology or economics, will, on the contrary, tend to publish in international journals structured around "common" international topics, in a manner similar to research in the NMS.

Finally, our results also clearly highlight the fact that the research output of SSH researchers from francophone universities is much more underestimated than the output of their colleagues from anglophone universities when only data from the WoS are used. In a context where global rankings of universities—for better or worse—are being increasingly produced and used,⁵ generally without caution, by university and government administrators, it is about time that we invest in developing new data sources or in improving existing ones to allow a fair evaluation of each and every university, irrespective of its disciplinary blend or language. Unfortunately, we are not optimistic about seeing such things happen, as it would mean admitting, for ranking producers, the important shortcomings of the various rankings they have produced over the years.

Acknowledgments

The authors thank Yves Gingras, Jean-Pierre Robitaille, Jillian Tomm, Matthew Wallace, and the two anonymous referees for useful comments on a previous version of this paper,

⁵See, for instance, the wiki page on college and university rankings: http://en.wikipedia.org/wiki/College_and_university_rankings

as well as Martin Boucher and Luc Grondin from the Centre d'édition numérique of the Université de Montréal for providing the source data of Érudit.

References

- Archambault, É., Vignola-Gagné, É., Côté, G., Larivière, V., & Gingras, Y. (2006). Benchmarking scientific output in the social sciences and humanities: The limits of existing databases. *Scientometrics*, 68(3), 329–342.
- Godin, B. (2002). The social sciences in Canada: What can we learn from bibliometrics? Project on the Measurement of the Social Sciences. Working Paper No. 1. Retrieved from: <http://www.csiic.ca/PDF/CSIIC.pdf>
- Hicks, D. (1999). The difficulty of achieving full coverage of international social science literature and the bibliometric consequences. *Scientometrics*, 44(2): 193–215.
- Hicks, D. (2004). The four literatures of social science. In H. Moed (Ed.), *Handbook of quantitative science and technology studies*. Dordrecht, NL: Kluwer Academic.
- Huang, M.H., & Chang, Y.W. (2008). Characteristics of research output in social sciences and Humanities: From a research evaluation perspective. *Journal of the American Society for Information Science and Technology*, 59(11), 1819–1828.
- Jin, B., & Wang, B. (1999). Chinese science citation database: Its construction and application. *Scientometrics*, 45(2), 325–332.
- Jin, B., Jiangong, Z., Chen, D., & Zhu, X. (2002). Development of Chinese scientometric indicators. *Scientometrics*, 54(1), 145–154.
- Larivière, V. (2010). A bibliometric analysis of Quebec's PhD students' contribution to the advancement of knowledge. Ph.D. Thesis, McGill University. Retrieved from: http://www.ost.uqam.ca/Portals/0/docs/Monographies/Thesis_Lariviere_Final.pdf
- Larivière, V., Archambault, É., Gingras, Y., & Vignola-Gagné, É. (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(8): 997–1004.
- Larivière, V., Macaluso, B., Archambault, É., & Gingras, Y. (2010). Which scientific elites? On the concentration of research funds, publications and citations. *Research Evaluation*, 19(1), 45–53.
- Liang, L.M. (2003). Evaluating China's research performance: How do SCI and Chinese indexes compare? *Interdisciplinary Science Reviews*, 28(1), 38–43.
- Meneghini R., Mugnaini R., & Packer A.L. (2006). International versus national oriented Brazilian scientific journals. A scientometric analysis based on SciELO and JCR-ISI databases. *Scientometrics*, 69(3), 529–538.
- Nederhof, A.J. (2006). Bibliometric monitoring of research performance in the social sciences and the humanities: A review. *Scientometrics*, 66(1), 81–100.
- Negishi, M., Sun, Y., & Shigi, K. (2004). Citation database for Japanese papers: A new bibliometric tool for Japanese academic society. *Scientometrics*, 60(3), 333–351.
- Tijssen R.J.W., Mouton, J., van Leeuwen, Th.N., & Boshoff, N. (2006). How relevant are local scholarly journals in global science? A case study of South Africa. *Research Evaluation*, 15(3), 163–174.
- White, H.D., Boell, S.K., Yu, H., Davis, M., Wilson, C.S., & Cole, F.T.H. (2009). Libcitations: A measure for comparative assessment of book publications in the humanities and social sciences. *Journal of the American Society for Information Science and Technology*, 60(6), 1083–1096.
- Zhou, P. Su, X., & Leydesdorff, L. (2010). A comparative study on communication structures of Chinese journals in the social sciences. *Journal of the American Society for Information Science and Technology*, 61(7), 1360–1376.

Appendix. Correlation between individual researchers' output indexed in the WoS and in Érudit, by discipline.

Discipline	R	Discipline	R
French/English	0.82	Electrical & Computer Engineering	–
Religious Studies & Vocations	0.71	Library & Information Sciences	–
Foreign Languages Literature, Linguistics & Area Studies	0.69	Mathematics	–
Earth & Ocean Sciences	0.53	Other Engineering	–
Planning & Architecture	0.27	Physics & Astronomy	–
Surgical Specialties	0.27	Mechanical & Industrial Engineering	–
Kinesiology / Physical Education	0.26	Resource Management & Forestry	–
Biology & Botany	0.21	Education	–0.01
Social Work	0.17	Economics	–0.03
History	0.15	Laboratory Medicine	–0.04
Philosophy	0.12	Agricultural & Food Sciences	–0.09
Medical Specialties	0.10	Rehabilitation Therapy	–0.09
Psychology	0.02	Law & Legal Studies	–0.11
Media & Communication Studies	0.02	Civil Engineering	–0.12
Geography	0.02	Other Social Sciences & Humanities	–0.13
Anthropology, Archaeology & Sociology	0.02	General Medicine	–0.13
Business	0.00	Political Science	–0.14
Chemical Engineering	–	Public Health & Health Administration	–0.20
Chemistry	–	Nursing	–0.24
Computer & Information Science	–	Fine & Performing Arts	–0.33
Dentistry	–	Other Health Sciences	–0.51