The Thesis Paradox:
An Empirical Study of the Impact of Doctoral Research

Vincent Larivière
Observatoire des sciences et des technologies, Centre interuniversitaire de recherche sur la science et la technologie, Université du Québec à Montréal, Case Postale 8888, Succursale Centre-Ville, Montréal (Québec), Canada, H3C 3P8, Tel +1.514.987.3000 #6799, Fax +1.514.987.8733. E-mail: lariviere.vincent@uqam.ca

Alesia Zuccala
School of Computing and Information Technology, University of Wolverhampton, MI Block, City Campus, Wulfruna Street, Wolverhampton, United Kingdom, WV1 1SB. E-mail: a.zuccala@wlv.ac.uk

Éric Archambault
Science-Metrix, 4572 avenue de Lorimier, Montréal, Québec, Canada, H2H 2B5 and Observatoire des sciences et des technologies (OST), Centre interuniversitaire de recherche sur la science et la technologie (CIRST), Université du Québec à Montréal, Montréal (Québec), Canada. E-mail: eric.archambault@science-metrix.com

Introduction

With an annual production of more than 41 000 Ph.D.s per year in the US alone (NSF, 2004), doctoral theses are expected to have a considerable impact on academic knowledge development. Previous research on dissertations examined topics ranging from quality and time-to-degree of a doctoral education (e.g., Ziolkowski, 1990; Bowen, Lord & Sosa, 1991; Gonzalaz, 1996; Katz, 1997), to the skills required of doctoral candidates (e.g., Isaac, Quinland & Walker, 1992; Barry, 1997), pre-thesis and post-thesis publication productivity (e.g., Lee, 2000; Anwar, 2004), and the role of dissertations as information sources (e.g. Boyer, 1973; Davidson, 1977). Despite these numerous studies on Ph.D.s and Ph.D. theses, there is a current lack of information concerning the scientific impact of this mode of diffusion.

This paper assesses the impact of theses based on their citation frequency in peer-reviewed papers and measures the evolution of this impact over time. Theses are becoming more available in electronic form (e.g., Networked Digital Library of Theses, 2005); hence, one can assume that they are increasingly consulted and used by researchers.

Methods

Drawing on data from the CD-ROM version of the *Science Citation Index*, *Social Sciences Citation Index* and *Arts and Humanities Citation Index*, this paper calculates the number of references made to theses in peer-reviewed papers between 1985 and 2004. For the National Sciences and Engineering (NSE), journals were assigned fields and subfields using the classification system developed by CHI Research (now ipIQ). For the Social Sciences and Humanities (SSH), a new classification system inspired by both CHI and Thomson Scientific was constructed using mutually exclusive fields. To calculate the number of references made to theses, an iterative retrieval process was carried out, starting with a keyword search for the term **thesis**, followed by an exclusionary search for false positives like *anaesthesis*. After several iterations, a simple solution emerged based on searching for *thesis** or **-thesis**. Although *thesis* is a common suffix, it is almost never used as a prefix. Sampling of a large number of results showed that the number of false positives was extremely small.

Results

Figure 1 shows that, starting in 1990, there has been a fairly steady increase in the number of citations made to theses. However, the number of papers in Thomson's databases has increased faster, consequently; the average number of citations to theses per paper has actually fallen.

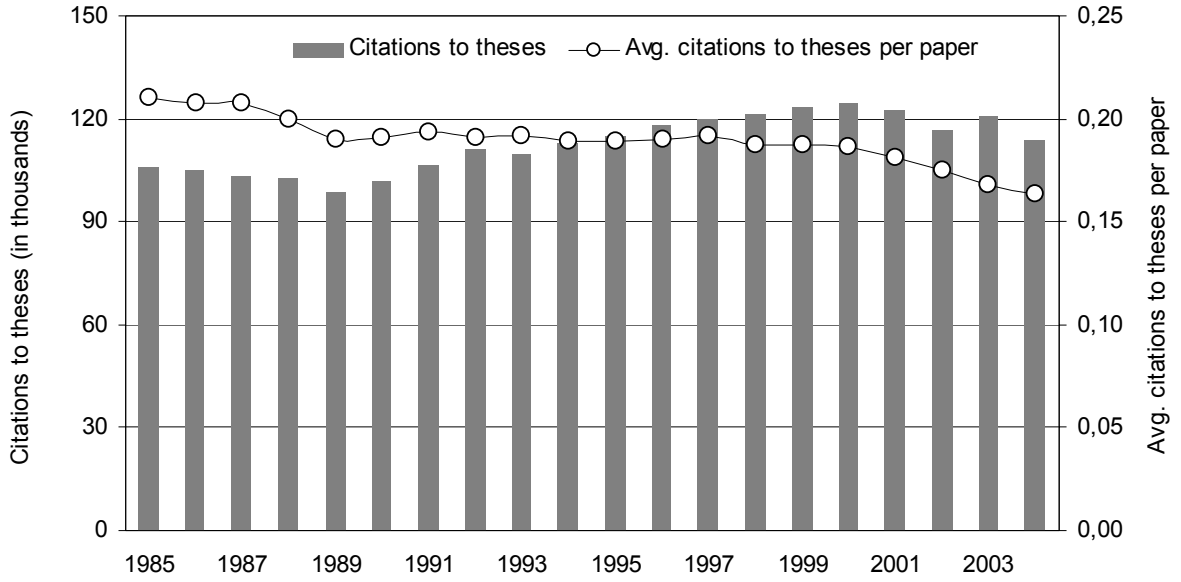

Figure 1. Citations to theses and average citations to theses per paper, 1985-2004

Figure 2 presents findings for the SSH while Figure 3 concentrates on the NSE. These data indicate that, in the SSH and with the exception of literature, there has been a marked decline in the share of references made to theses in peer-reviewed papers. On average, theses from the SSH received only 1.2% of the references in 1985 and this fell by 0.5 percentage point during the 20-year period.

In the NSE, we see a similar decline in thesis citations. The first health sciences cluster in Figure 2 shows that references to theses are almost non-existent. This may be due to the importance of published papers in these fields, or to a common tendency for students to obtain a doctoral degree from cumulating published papers. In chemistry, physics and psychology, there is an intermediate cluster for the proportion of references made to theses, and in the third cluster lead by engineering and technology, more weight is given to the use of theses in published research, yet the decline in citations over time is still evident.

Further analyses of our results pointed to another interesting finding: theses are several times more likely to be self-cited than scientific production in general. While, over the 20-year period, 10% of all references were self-citations, more than 25% of theses were self-cited.

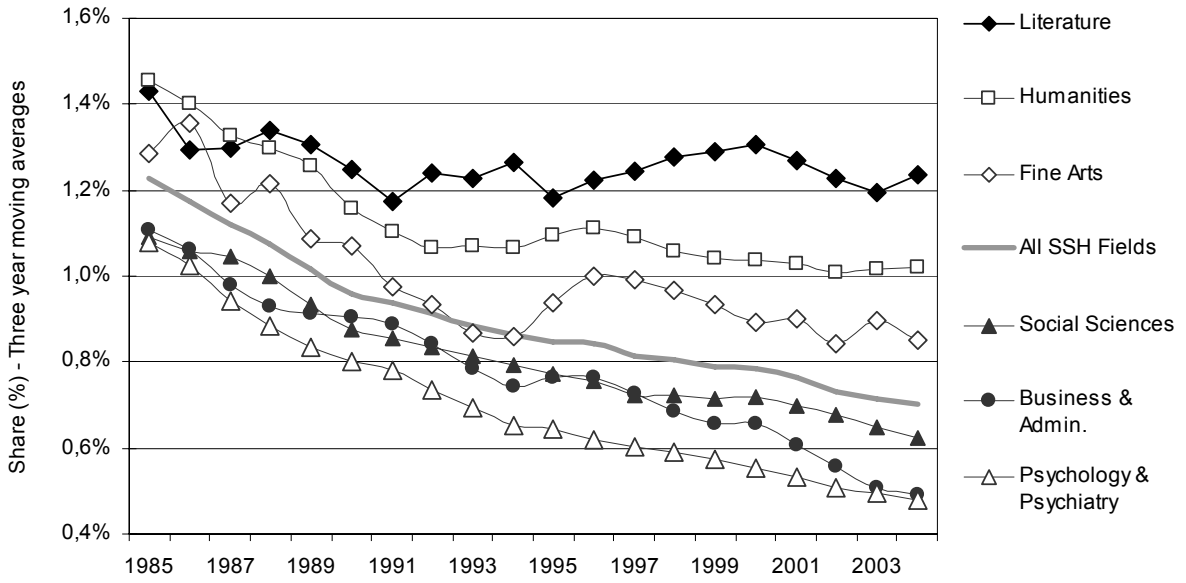

Figure 2. Share of Citations Made to Theses in the SSH, 1985-2004

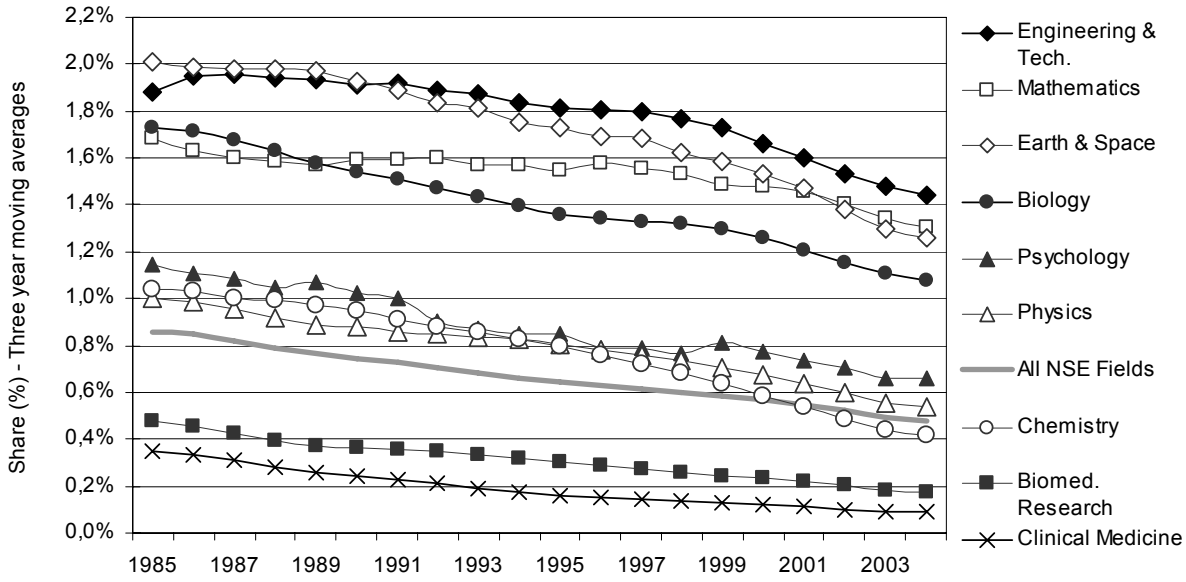

Figure 3. Share of Citations Made to Theses in the NSE, 1985-2004

Conclusion

The results of this study present a paradox: more theses are produced, and are more accessible to scholars in electronic form, yet their scientific impact seems to be declining. This does not mean that a doctoral dissertation is a poor source of scholarly information. New knowledge is percolating the academic system, thus it may be that scholars prefer to cite published papers and books *derived* from graduate research, rather than actual theses. For some researchers it is potentially easy to overlook the availability of theses as sources of information, given that so many other types of publications (journal articles, research reports, etc.) are also available on the Web. From a science policy point of view, more consideration needs to be given to the development of thesis repositories. If scholars wish to maximize the readership of their research, mainly their dissertation research, newly created thesis repositories should be better marketed.

Bibliography

Anwar, M. A. (2004). From doctoral dissertation to publication. A study of 1995 American graduates in Library and Information Science. *Journal of Librarianship and Information Science, 36*, 151-157.

Barry, C.A. (1997). Information skills for an electronic world: training doctoral research students. *Journal of Information Science 23(3)*, 225-238.

Bowen, W. G., Lord, G. & Sosa, J. A. (1991). Measuring time to the doctorate: Reinterpretation of the evidence. *Proceedings of the National Academy of Sciences of the United States of America, 88(3)*, 713-717.

Boyer, C. (1973). *The doctoral dissertation as an information source: a study of scientific information flow*. Metuchen, NJ: Scarecrow Press

Davinson, D. E. (1977). *Theses and dissertations as information sources*. Hamdon, CT. Linnet Books.

Gonzalez, A. (1996). Graduation rates and times to completion for doctoral programs in Canada. *Education Quarterly Review, 3(2)*, 45-56.

Isaac, P. D., Quinlan, S. V., & Walker, M. M. (1992). Faculty Perceptions of the Doctoral Dissertation. *Journal of Higher Education, 63(3)*, 241-268.

Katz, E. L. (1997). Key players in the dissertation process. *New Directions for Higher Education, 99*, 5-16.

National Science Foundation, Division of Science Resources Statistics (2004), *Science and Engineering Doctorate Awards: 2003*, NSF 05-300. Available at: http://www.nsf.gov/statistics/nsf05300/pdfstart.htm

Networked Digital Library of Theses and Dissertations. (2005). Available at: http://www.ndltd.org/about.en.html.

Lee, Wade M. (2000). Publication trends of doctoral students in three fields from 1965-1995. *Journal of the American Society for Information Science 51(2)*, 139-44.

Ziolkowski, T. (1990). The Ph.D. Squid. *The American Scholar, 59(2)*, 177-195.