

The declining scientific impact of theses: Implications for electronic thesis and dissertation repositories and graduate studies

VINCENT LARIVIÈRE,^{a,d} ALESIA ZUCCALA,^b ÉRIC ARCHAMBAULT^{a,c}

^a *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)*

^b *Rathenau Instituut, Den Haag (The Netherlands)*

^c *Science-Metrix, Montréal, Québec (Canada)*

^d *Graduate School of Library and Information Studies, McGill University, Montréal, Québec (Canada)*

Although the writing of a thesis is a very important step for scientists undertaking a career in research, little information exists on the impact of theses as a source of scientific information. Knowing the impact of theses is relevant not only for students undertaking graduate studies, but also for the building of repositories of electronic theses and dissertations (ETD) and the substantial investment this involves. This paper shows that the impact of theses as information sources has been generally declining over the last century, apart from during the period of the 'golden years' of research, 1945 to 1975. There is no evidence of ETDs having a positive impact; on the contrary, since their introduction the impact of theses has actually declined more rapidly. This raises questions about the justification for ETDs and the appropriateness of writing monograph style theses as opposed to publication of a series of peer-reviewed papers as the requirement for fulfilment of graduate studies.

Introduction

The writing of a Ph.D. thesis (see footnote 1 on next page) is mandatory for scientists seeking to pursue a career in research. By writing a Ph.D. thesis and contributing to the advancement of scientific knowledge in a research field, a researcher receives recognition from those who will become his peers confirms the new

Received February 12, 2007

Address for correspondence:

VINCENT LARIVIÈRE

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 (Québec) H3C 3P8, Canada

E-mail: lariviere.vincent@uqam.ca

0138–9130/US \$ 20.00

Copyright © 2007 Akadémiai Kiadó, Budapest

All rights reserved

researcher's status as a member of the chosen scientific field [BOURDIEU, 1975]. Though the doctorate has existed since the Middle Ages, it was only at the beginning of the 19th century that the Ph.D. degree became a diploma associated to the production of original scientific research and the training of new researchers when the Prussian Minister for Education, Wilhelm von Humboldt, established a new university model at the University of Berlin. The Humboldtian university model, which institutionalized scientific research within the walls of the university, with professors having to both teach and perform research, became the standard in other countries – such as Canada and the United States – before WWI [GINGRAS, 2003].

Since then, the Ph.D. degree's legitimacy as an entry point to the scientific community has remained relatively unchallenged. The main requirement for obtaining a Ph.D. degree is to produce a thesis that makes an "original" and "significant contribution" to the student's subject area (see, e.g., [COUNCIL OF GRADUATE SCHOOLS, 1997; TINKLER & JACKSON, 2000; WINTER & AL., 2000]). However, its form has undergone some changes over the last 25 years, especially in natural science and engineering (NSE) programmes. While in the social sciences and humanities (SSH) the required form of the thesis has not changed substantially, theses in the NSE are increasingly consisting of a series of published articles. There is no immediately available information on the number of Ph.D. theses that are being taking the form of a compilation of publications, but a quick search of the thesis submission guidelines from different graduate schools and departments shows that a high percentage will allow this type of submission. Also, like other research results, Ph.D. theses are becoming increasingly available in electronic form [WEISSER & WALKER, 1997]. Electronic modes of diffusion include initiatives taken by Ph.D. graduates to publish their theses on personal websites, electronic archives set up by institutions² and broader initiatives that allow international users to deposit their theses, and to search for theses written by students worldwide (e.g., [NETWORKED DIGITAL LIBRARY OF THESES AND DISSERTATIONS – NDLTD, 2005]).

This paper has three objectives. First it aims to analyze the overall scientific impact – as measured by citations – of theses on new knowledge. Given an annual production of more than 41,000 Ph.D.s per year in the U.S. alone [NATIONAL SCIENCE FOUNDATION. DIVISION OF SCIENCE RESOURCES STATISTICS, 2004], one might expect that doctoral theses would have a considerable impact on the development of scientific knowledge. Second, we want to investigate whether the increased availability of theses is increasing their scientific impact. Because of their wider availability (digital theses can be accessed online) and the reduced transaction costs associated with preparing and

¹ There is no standardized definition that clearly distinguishes between a thesis and a dissertation. If one trusts Google, the term thesis is unequivocally more popular than the term dissertation at both the master's and doctoral levels. In this paper, thesis refers to master's and doctoral level theses and dissertations.

² See : e.g. the University of Nottingham e-thesis archive: <http://etheses.nottingham.ac.uk/information.html>

obtaining readable copies of theses, citation frequency for the most recent years (i.e., between 2000 and 2004) should be substantially higher than, for instance, before 2000. Thirdly, using historical bibliometric data, this paper examines long term trends that might provide clues to changes in thesis citation rates.

Background

Previous research on dissertations has examined such topics as quality and time-to-degree of a doctoral education (e.g., [ZIOLKOWSKI, 1990; BOWEN & AL., 1991; GONZALEZ, 1996; KATZ, 1997]), skills required of doctoral candidates (e.g., [ISAAC & AL., 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 the fairly large number of studies on Ph.D.s and Ph.D. theses, there is a relative scarcity of information concerning the scientific impact of this mode of knowledge diffusion. One of the aspects that has been examined is the value of doctoral theses as a research collection development tool [EDWARDS, 1999; BEILE & AL., 2004], or tool for measuring a faculty's use of library journals [ZIPP, 1996]. BEILE & AL. [2004] argue that it is not always appropriate to use a thesis to evaluate a library collection, particularly if the emphasis is on the evaluation outcome rather than the actual quality of the thesis. Some researchers have looked at how doctoral students use information in the development of their theses, in terms of both the most highly cited material and the currency of the literature and the use of specific journals (e.g., [WALCOTT, 1991; BUTTLAR, 1999; GOODEN, 2001; KUSHKOWSKI & AL., 2003]). A common finding is that journal articles tend to be cited more often than monographs, and that theses account for a very small proportion of the cited literature overall. KUSHKOWSKI & AL. [2003] discovered that – regardless of disciplinary affiliation – graduate students favoured current research in their bibliographies, and that the number of references in theses varied by discipline. More recently, KUSHKOWSKI [2005] investigated the Web citation behaviour of the authors of print theses versus that of the authors of electronic theses and found that Web citations account for 2.2% of the citations in print and 5.4% of the citations in ETDs. The persistence of Web citations in theses has, so far, been uniformly poor.

For members of a scholarly community to read, evaluate and potentially cite a thesis, there is a need for it to be widely disseminated or at least made readily accessible. DAVINSON'S [1977] text on *Theses and Dissertations as Information Sources* recalls the past practice of “exchanging theses” and the requirement of doctoral students to have them “printed and bound up in book form at their own expense” (p. 60). Throughout the 1960s, this exchange practice declined (e.g., [ASH, 1969]); bound theses or microfilmed copies began to be stored at the institutions from which students graduated and were made available through interlibrary loans. Loan practices

varied: “many British university libraries have been remarkably generous in their loan policies... others have been sometimes less generous” (p. 62). DAVINSON [1977] and REPP & GLAVIANO [1987] claim that availability in North America (and subsequently in Europe) was affected by the fact that “the presence of *Dissertation Abstracts International (DAI)* has solved many of the problems of access” [DAVINSON, 1977 : 65]. DAVINSON [1977 : 66] was critical of this service, stating that it is:

little more than a publishers’ catalogue with each of the entries accompanied by a price tag. ... Would-be users are often obliged to buy copies of material they think may be of value to them, judging [only] from the abstract. They may be disappointed (and substantially out of pocket) when they receive it.

Currently, there is a significant movement towards the digitization of theses for open access at both local (for instance, through the development of institutional e-thesis repositories) and international levels, through the NDLTD [2005].³ The NDLTD was established at Virginia Tech with funding provided by the U.S. Department of Education. Several universities around the world have become official contributing members of this open access initiative, and many others are expressing interest in doing so, or making arrangements to participate [WEISSER & WALKER, 1997]. There are now more than 200,000 documents in the NDLTD repository.

Methods

Our analysis is based on Thomson Scientific’s bibliographic databases. Data from 1900 to 1944 are drawn from the *Century of Science*, which indexes 266 journals covering most natural sciences and medical fields. 1945 to 1979 data are from the Web of Science, and data for 1980 to 2004 are from the CD-ROM versions of the *Science Citation Index (SCI)*, *Social Sciences Citation Index (SSCI)* and *Arts and Humanities Citation Index (AHCI)*. Thomson Scientific databases index several types of documents as source items, among which only articles, research notes and review articles are generally used for bibliometric studies. Since they are the main media for diffusing new knowledge, this paper focuses on these three document types.

Even though cited references in the SCI, SSCI and AHCI are not “tagged” with a document type, regular patterns in cited references facilitate the identification of several types of cited documents. For instance, MOED [2005] and LARIVIÈRE & AL. [2006] showed that it was possible to isolate among cited documents, references to journal articles. A similar approach is used here. To identify references to theses from all cited material, an iterative retrieval process was carried out, starting with a keyword search for the term *thesis*, followed by an exclusionary search for false positives. Taking

³ See <http://www.ndltd.org/about.en.html>

into account that most of “noise” was generated from documents not starting with the string *thesis* (hypothesis, synthesis, anaesthesia, antithesis, metathesis, etc.), it was found that searches using only the string *thesis** at the beginning of the document were very accurate in terms of isolating references made to theses. Although *thesis* is a common suffix, it almost never occurs as a prefix. The sampling of a large number of results showed that the number of false positives was extremely small. Moreover, our extensive validation showed that *thesis* appeared to be the expression used by Thomson Scientific to index both theses and dissertations.

This paper uses the field and sub-field classifications employed by the U.S. National Science Foundation in its Science and Engineering Indicators. The main advantage of this classification over that used by Thomson Scientific is that fields are mutually exclusive and, therefore, each journal appears only once. Hence, this classification overcomes problems associated with multiple counts. Unfortunately, there is no similar classification for journals in the SSH. Consequently, a new classification system, inspired by those used by NSF and Thomson Scientific, was constructed to classify SSH journals and associate them with mutually exclusive fields and sub-fields.

Findings

Figure 1 presents the total number of references made by papers in Thomson Scientific’s databases between 1900 and 2004 as well as the number of citations to theses during these years. The impact of the two World Wars on citation rates is immediately obvious, but what is more important for this paper is that, whereas the number of references continues to rise unabated, the number of citations to theses has been levelling off since about 1980.

Figure 2 shows that there have been three main trends in citations to theses. The first movement, ending with World War II, clearly reveals the diminishing importance of theses in overall citation traffic. This was followed by the period that could be termed the “golden years” of academic research, during which there was a sustained growth in public spending on scientific research accompanied by a growing number of students attending and graduating from universities and, subsequently, finding jobs as researchers in universities, public administrations and private firms. The last period starts in the mid 1970s and is either a return to the trend observed before the golden years or a new downturn in the importance of theses as citeable material.

Drawing on data covering the last 25 years of scientific activity, Table 1 shows that theses account for only a small proportion of cited documents: about 1% in the SSH and about two-thirds of 1% in the NSE. Nevertheless, on average, between 1980 and 2004, about one paper in four in the SSH and one in six in the NSE cited a thesis.

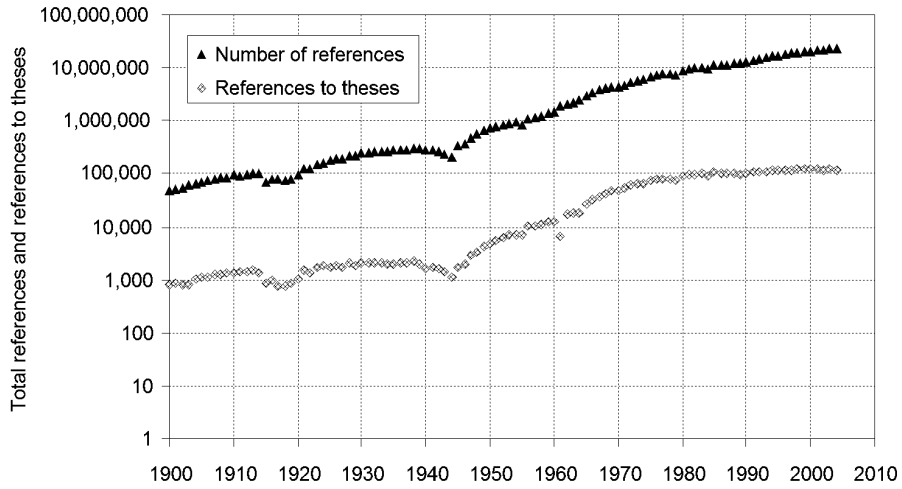


Figure 1. Total number of references and references to theses, 1900–2004

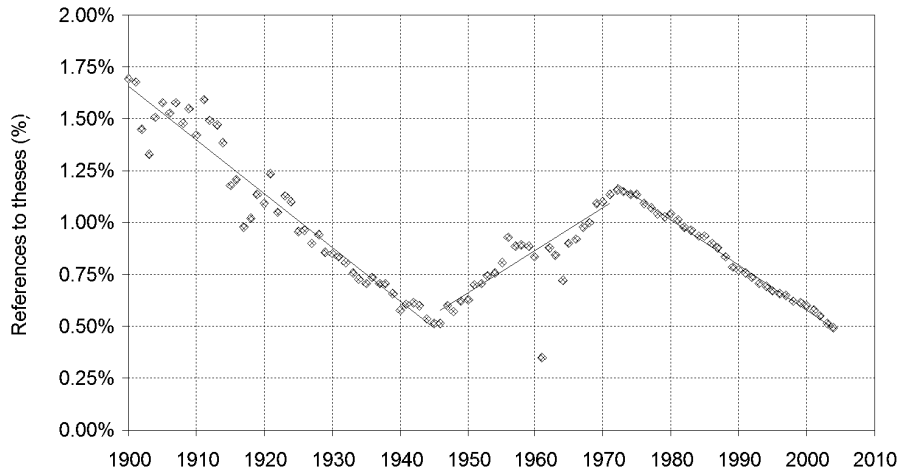


Figure 2. Share of references to theses, 1900–2004

Table 1. References to theses, 1980–2004

Indicator	Social Sciences and Humanities	Natural Sciences and Engineering
N references the theses	513,745	2,214,338
N articles	1,885,828	12,350,777
N references	55,364,666	329,304,208
% references to theses	0.93	0.67
N references to theses per paper	0.27	0.18

Figure 3 is useful for examining the hypothesis that electronic thesis repositories would help to increase the impact of theses. It can be seen that the absolute number of references to theses started to decline in 2001, and that the decline in the average number of references to theses began to accelerate in 2001. This shows that electronic theses repositories, at best, have merely slowed down the downward trend. Figure 3 shows that there is a long-term declining trend in the number of references to theses per paper: whereas in 1980 there were about 0.21 citations to theses by papers, this number has fallen over the last 25 years, resulting in an average of 0.16 references to theses by papers in 2004.

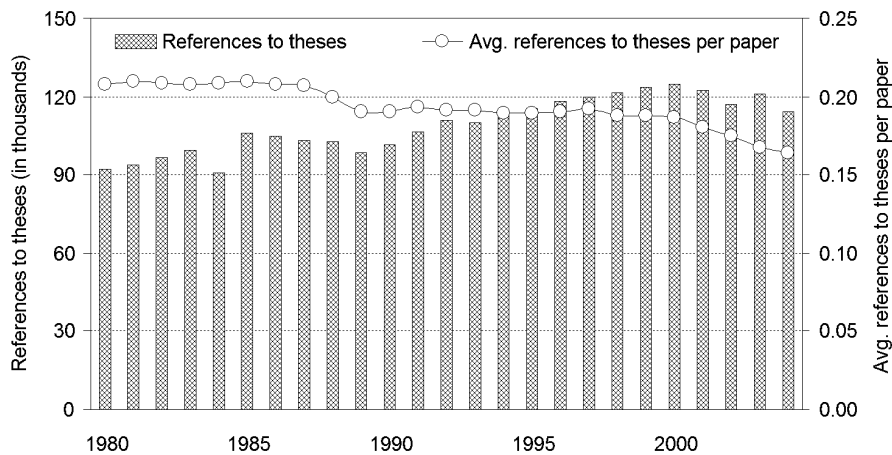


Figure 3. References to theses and average number of references to theses per paper, 1980-2004

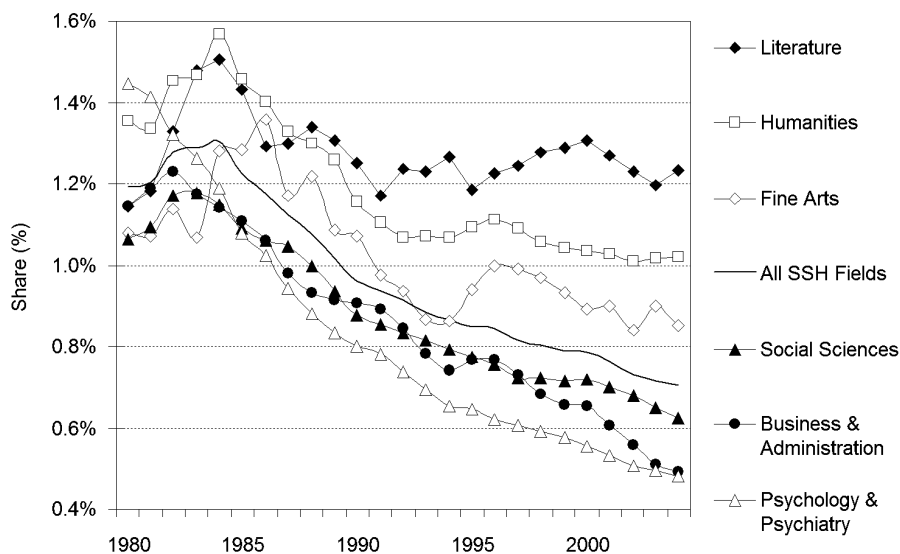


Figure 4. Share of references made to theses in the social sciences and humanities, 1980–2004 (three-year moving averages)

Figure 4 presents findings for the SSH; Figure 5 provides data for the NSE. The SSH data indicate that, with the exception of literature, there has been a marked decline in the share of references to theses in peer-reviewed papers. On average, SSH theses received only 1.2% of the citations made in 1985; over the 20 year period to 2004 this figure dropped by 0.5 percentage points.

The NSE data show a similar decline. The first health sciences cluster in Figure 5 shows that here there were almost no citations made to theses. This may be due to the importance of published papers in these fields, or to the growing tendency towards an accumulation of published papers rather than a thesis for fulfilment of doctoral degree requirements. There is an intermediate cluster for chemistry, physics and psychology for the proportion of references made to theses. A third cluster, led by engineering and technology, shows more importance to citations to theses in published research. Nevertheless, the decline in citations over time is evident.

The analysis depicted in Figure 6 shows that citations to theses are several times more likely to be self-citations than citations to other scientific production. Over the 25-year period, 10% of all references in the NSE were self-citations, but more than 25% of NSE citations to theses were self-citations. Likewise, 5% of citations in the SSH were self-citations, but more than 15% of citations to theses were self-citations.

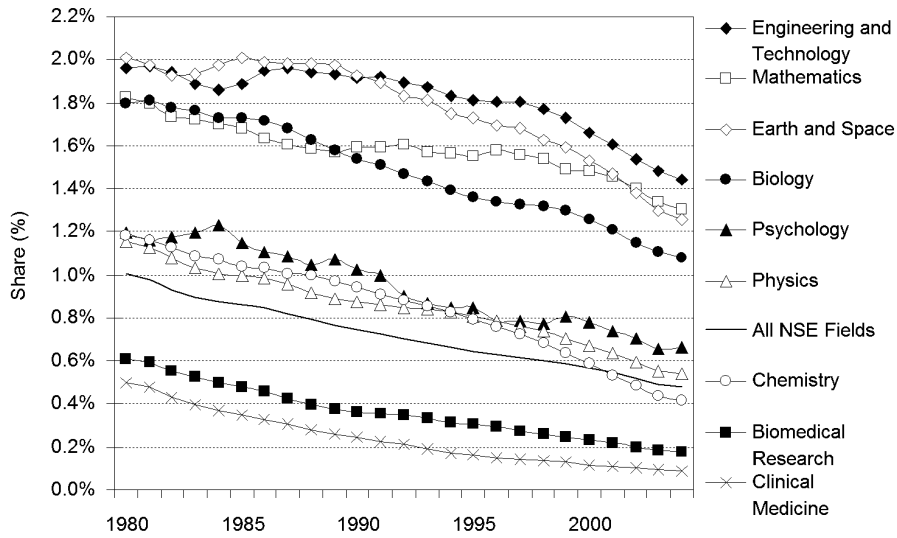


Figure 5. Share of references made to theses in the natural sciences and engineering, 1980–2004 (three-year moving averages)

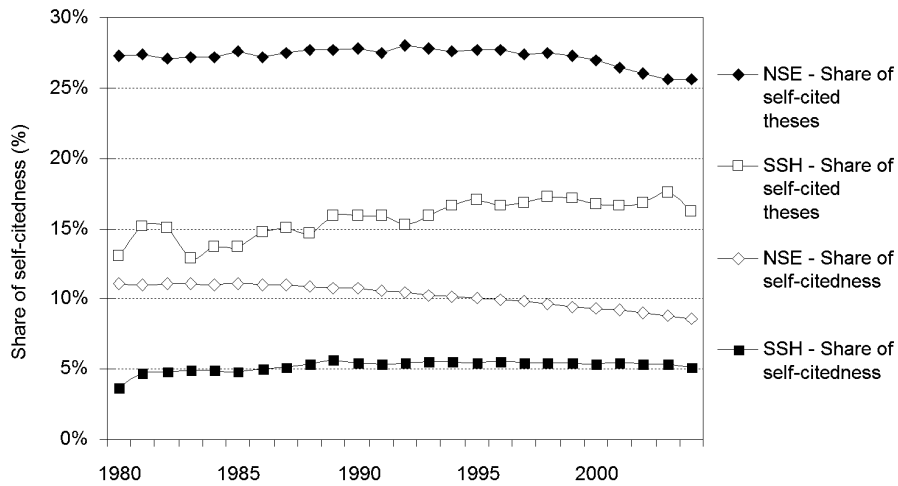


Figure 6. Share of self-citations made to theses and average self-citedness, NSE and SSH, 1980–2004

Discussion

The results of this study are paradoxical: while theses are becoming increasingly accessible to scholars in electronic format, their overall scientific impact seems to be declining. This does not necessarily mean that a doctoral thesis is a poor source of scholarly information. It may be the case that as new knowledge is permeating the academic system, scholars prefer to cite published papers and books derived from graduate research, rather than the original theses. In some cases, it may be that the availability of theses as sources of information has been overlooked, given the increasing amount of other types of publication (journal articles, research reports, etc.) that are being made available on the Web.

From a policy point of view, more consideration needs to be given to the development of e-thesis repositories, which, if scholars wish to maximize the readership of theses, may need to be promoted. However, their promotion will not guarantee a reversal of the observed trend towards a marginalization of theses as a source of scientific information. In fact, the data presented in this paper show that if development of e-thesis repositories were to be based on the use of theses as sources of scientific information, there would be little justification for it. There is an argument for their development based on the intrinsic requirement of a thesis or dissertation, i.e. that it should be an original piece of work. Currently, it is very difficult, if not impossible, for students undertaking postgraduate studies to be confident that no other student around the world has covered a particular niche of research. In this respect, e-thesis repositories would play a very important role in allowing new researchers to search for and examine examples of “prior art”, and it is likely, although not by any means certain, that this would increase the number of references to theses made by graduate students.

Although the data in this paper point to the marginal role played by theses as sources of scientific information, they do not identify a possible causal mechanism. Two explanations can be proposed. First, the peaks and troughs in the graph in Figure 2 (overall share of references to theses) may be linked to supply and demand. The more theses that are produced and the higher their percentage in total scientific contributions, the more frequently they will be cited. Figure 7 presents only partial data on the “supply side”, but these suggest that there is a relationship between the number of Ph.D. graduates [NATIONAL SCIENCE FOUNDATION, DIVISION OF SCIENCE RESOURCES STATISTICS, 2006] and the number of citations to theses. For instance, the number of graduates in the U.S. peaked around 1973, which more or less coincides with the period when the share of references to theses peaked (see Figure 7 and Figure 2). Figure 7 also shows that citations to theses levelled off a few years after this 1973 peak in graduate numbers in the U.S.

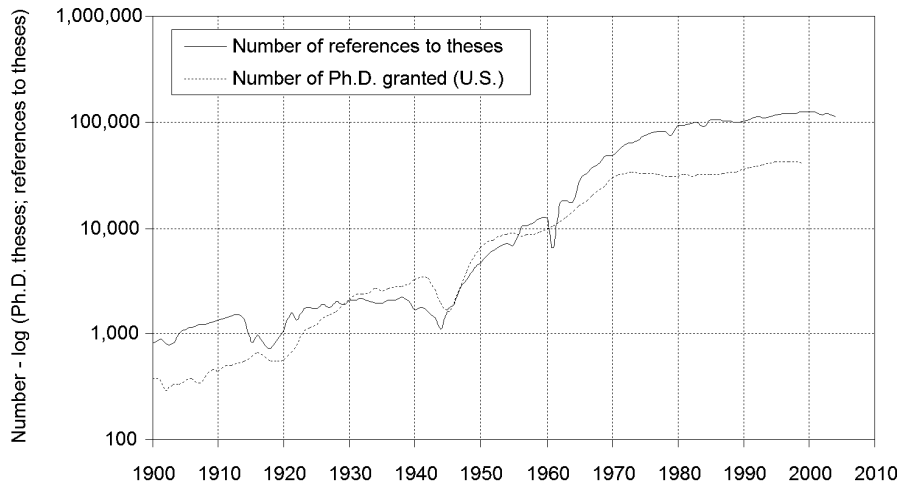


Figure 7. Number of references to theses and Ph.D.s granted, 1900–2004

A second explanation for the drop in the proportion of citations made to theses might be that, starting in the mid 1970s, the popularity of series of published papers compiled as a thesis, rather than the traditional monograph, has grown. This would add to the observed shift in citation patterns, with fewer citations to theses but more to graduate study-related papers.

Overall, this paper raises important questions concerning the modus operandi of graduate programmes. It would be legitimate to ask why graduate students are still encouraged in many fields to produce monograph-style theses if their scientific impact is diminishing. The monograph-type thesis is still relevant as a significant test of a student's knowledge, research skills and effectiveness at communicating. On the other hand, graduation-by-papers provides students with the opportunity to practice research and improve their writing skills, both of which are important in terms of the dominant form of scientific contribution in the majority of scientific fields [LARIVIÈRE & AL., 2006]. A papers-based graduate approach might be a more effective training for researchers if publishing papers is what will concern them during most of the rest of their careers.

*

The authors wish to thank François Vallières and Gilles Renaud for the construction of the bibliometric databases, and Yves Gingras, Jean-Pierre Robitaille and Jean Lebel for their detailed and insightful comments on an earlier version of this paper.

This paper is a revised and extended version of the paper entitled "The thesis paradox: an empirical investigation on the impact of doctoral research" that was originally presented at the 9th International Conference on Science and Technology Indicators, held in Leuven, Belgium, in September 2006.

References

- 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.
- ASH, J. (1969). The exchange of academic dissertations. *College and Research Libraries*, 30 (3) : 237–241.
- BARRY, C. A. (1997). Information skills for an electronic world: training doctoral research students. *Journal of Information Science*, 23 (3) : 225–238.
- BEILE, P. M., BOOTE, D. N., KILLINGSWORTH, E. K. (2004). A microscope or a mirror? A question of study validity regarding the use of dissertation citation analysis for evaluating research collections. *Journal of Academic Librarianship*, 30 (5) : 347–353.
- BOURDIEU, P. (1975). La spécificité du champ scientifique et les conditions sociales du progrès de la raison. *Sociologie et sociétés*, 7 (1) : 91–118.
- 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*. Scarecrow Press, Metuchen, NJ.
- BUTTLAR, L. (1999). Information sources in library and information science doctoral research. *Library & Information Science Research*, 21 (2) : 227–245.
- COUNCIL OF GRADUATE SCHOOLS (1997). The role and nature of the doctoral dissertation: a policy statement. Retrieved September 20th, 2005 from: <http://www.cgsnet.org/PublicationsPolicyRes/role.htm>.
- DAVINSON, D. E. (1977). *Theses and dissertations as information sources*. Linnet Books, Hamdon, CT.
- EDWARDS, S. (1999). Citation analysis as a collection development tool: a bibliometric study of polymer science theses and dissertations. *Serials Review*, 25 (1) : 11–20.
- GINGRAS, Y. (2003). Idées d'universités : enseignement, recherche et innovation. *Actes de la Recherche en Sciences Sociales*, 148 : 3–7.
- GONZALEZ, A. (1996). Graduation rates and times to completion for doctoral programs in Canada. *Education Quarterly Review*, 3 (2) : 45–56.
- GOODEN, A. M. (2001). Citation analysis of chemistry doctoral dissertations: An Ohio State University case study. *Issues in Science and Technology Librarianship*, Available at: <http://www.istl.org/01-fall/refereed.html>
- 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.
- KUSHKOWSKI, J. D. (2005). Web citation by graduate students: A comparison of print and electronic theses. *portal: Libraries and the Academy*, 5 (2) : 259–276.
- KUSHKOWSKI, J. D., PARSONS, K. A., WIESE, W. H. (2003). Master's and doctoral thesis citations: Analysis and trends of a longitudinal study. *portal: Libraries and the Academy*, 3 (3) : 459–479.
- LARIVIÈRE, V., ARCHAMBAULT, E., GINGRAS, Y., VIGNOLA-GAGNÉ, E. (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.
- LEE, W. 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.
- MOED, H. F. (2005). *Citation Analysis in Research Evaluation*. Dordrecht: Springer.
- 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>

- NATIONAL SCIENCE FOUNDATION, DIVISION OF SCIENCE RESOURCES STATISTICS (2006). *U.S. Doctorates in the 20th Century*. NSF 06-319. Available at: <http://www.nsf.gov/statistics/nsf06319/start.cfm>
- NETWORKED DIGITAL LIBRARY OF THESES AND DISSERTATIONS (NDLTD) (2005). Retrieved August 22, 2005 from <http://www.ndltd.org/about.en.html>.
- REPP, M. J., GLAVIANO, C. (1987). Dissertations: A Study of the Scholar's Approach. *College & Research Libraries*, 48(2) : 148–159.
- TINKLER, P., JACKSON, C. (2000). Examining the doctorate: Institutional policy and the Ph.D. examination process in Britain. *Studies in Higher Education*, 25 (2) : 167–180.
- WALCOTT, R. (1991). Characteristics of citations in geoscience doctoral dissertations accepted at United States academic institutions 1981-1985. *Science & Technology Libraries*, 12 (2) : 5–16.
- WEISSER, C. R., WALKER, J. R. (1997). Electronic theses and dissertations: digitizing scholarship for its own sake. *The Journal of Electronic Publishing*, 3 (2). Retrieved October 14, 2005 from <http://www.press.umich.edu/jep/03-02/etd.html>.
- WINTER, R., GRIFFITHS, M., GREEN, K. (2000). The “academic qualities” of practice: What are the criteria for a practised-based Ph.D.? *Studies in Higher Education*, 25 (2) : 25–37.
- ZIOLKOWSKI, T. (1990). The Ph.D. Squid. *The American Scholar*, 59 (2) : 177–195.
- ZIPP, L. S. (1996). Thesis and dissertation citations as indicators of faculty research use of university library journal collections. *Library Resources and Technical Services*, 40 (4) : 335–342.