How Influential is Brooks' Law? A Citation Context Analysis of Frederick Brooks' *The Mythical Man-Month*

Katherine W. McCain* and Laura J. Salvucci**

*Kate.McCain@cis.drexel.edu
College of Information Science & Technology, Drexel University,
3141 Chestnut Street, Philadelphia, PA 19104 (USA)

*Laura.Salvucci@villanova.edu
Department of History, Villanova University, 800 Lancaster Ave.
Villanova, PA 19085, (USA)

Abstract
Five hundred seventy-four citation contexts from 497 journal articles citing an edition of Frederick P. Brooks, Jr.'s *The Mythical Man-Month* (*MMM)* were content-analyzed to assess the degree to which *MMM* has become a "concept symbol" in the sense of Small, 1978. Data were collected for the period 1975-May, 1999. A classification scheme comprising 15 content classes in four areas (generalia, project management issues, building the system, other) was developed and each context uniquely assigned to one class. Citing article contexts were assigned a subject code based on the publishing journal and its ISI subject and LC class assignments. *MMM* represents a variety of different concepts with differing emphasis depending on the subject area of the citing article. The eponymous concept "Brooks' Law" (the "mythical man-month" or "adding more people to a late project makes it later"), although the most prominent concept across the board, accounted for less than one quarter of the 513 contexts assigned to one of the first three content class groups. *MMM* is also frequently cited in Software Engineering, Computer Science, and Information Systems, as a landmark work in software engineering, while project management issues are the second most frequent content class in Management.

Introduction

In the almost 30 years since its publication, Frederick P. Brooks, Jr.'s book, *The Mythical Man-Month (MMM)* (Brooks, 1975), has become established as a landmark publication in the area of software project management. *MMM* is a collection of essays on issues encountered and lessons learned during the development of IBM's OS360—the first large-scale computer operating system. It was the first major work that discussed the problems in managing large software development projects. An anniversary edition (Brooks, 1995) added several additional essays, including a second influential publication "No Silver Bullet" (previously published as Brooks, 1987). The title of both editions refers to the eponymous "Brooks' Law" – "adding more people to a late software project makes it later" — but, as Verner et al (1999) note, the book contains many other cautions for project managers as well.

In their paper, Verner et al. (1999) briefly discuss the general quantitative impact of *MMM* and the geographic breadth and subject span of the citing articles as preface to a survey of current software project management practices. In the present paper, we explore more systematically the changing influence of *MMM* over its first quarter century through a citation context analysis of the ISI-indexed articles citing *MMM* in any of its various editions.

In the use of citation context analysis, we build on the work of Henry Small—specifically two papers—Small (1978) and Small(1982). In his 1978 paper, Small suggested that, by citing a document and embedding the reference in textual commentary, authors impart meaning to these cited works which then, over time, become symbols of the concepts they contain (or which authors attach to them). The nature and variety of these concepts can be studied through a content analysis of the citation context (Small, 1982) and, over a set of citing documents, the percent uniformity (the degree to which authors demonstrate consensus on the nature of the cited concept) can be calculated.

While Small's 1978 and 1982 papers are highly cited, only a few studies have actually used citation context analysis to characterize the concept symbol nature of cited works by examining the content of the citation context. Small (1978) and Small & Greenlee (1980) examined the level of consensus in

Proceedings of ISSI 2005, pages 327–334.
Edited by P. Ingwersen and B. Larsen.

documents in chemistry and recombinant DNA, respectively. Cozzens (1985) assessed the level of consensus concerning two papers, one in neuropharmacology and one in sociology of science as did Hargens (2000) for papers in economics and cognitive psychology. McCain & Turner (1989) included a content analysis of citation contexts as part of a classification scheme intended to account for different citation histories in molecular genetics. Three studies have looked at individual monographs. Garfield (1985a, reprinted in 1985b) and Furner (2003) both provide in-depth discussions of the important concepts and citation history of Price's *Little Science, Big Science,* while Lewison (2004) compares the citation history of James Bond's ornithological monographs to a selection of similar works by other authors.

There have been very few bibliometric studies of software engineering (SE). Coulter et al. (1998) published a co-word analysis of SE based on the ACM Classification scheme and identified major research themes. Verner et al (1999, 2001) discuss current practices in software project management and visibility of software engineering authors. Marion & McCain (2001) map software engineering journals and McCain et al. (2003) compare maps based on author co-citations and card sorting.

Methods

Citing articles and contexts

We searched the ISI files (Dialog files 434, 34, 7, 439) for all citations to the various editions of *MMM* as a cited work. We obtained as many articles as possible and examined each for citations to *MMM* embedded in the text. (We processed only English-language articles, which were the overwhelming majority retrieved.) For each article in our database that had an explicit embedded citation, we transcribed the citing context (sentences surrounding the embedded citations) and entered it, along with the article-level information, in a second database The outcome was the creation of two databases, one of 497 citing articles containing at least one embedded reference to *MMM* and another of 574 records of individual citation contexts and their associated article and journal information. Most articles contained a single citation instance, but a noticeable number had 2 or more.

Journal subject areas

We needed to assign a single, useful subject class to each journal title, to be inherited by all articles included in the study and, ultimately, all contexts in those articles. The final subject classification was based primarily on an assessment of both the ISI subject classes and the Library of Congress class for each title (from *Ulrich's International Periodicals Directory* or from the catalog of a holding library). We also classified as "software engineering" all journals identified as core SE journals by Marion & McCain (2001) and identified as "information systems" journals, those journals with the phrase in their titles or the ISI subject class title.

Concept Classification Analysis

Small (1982) describes two different uses of content analysis in the study of citation contexts. The first is a classification of the "types or functions of references in scholarly texts." The second uses content analysis to identify the concepts "attributed to the cited work by the citing author." While Small points out that content issues may be conflated with functional considerations in individual classification schemes, the explicit analysis of the citation content in the text surrounding the embedded citation (the citation context) allows the identification of the idea symbolized by the cited work. The uniformity of textual content can be read as a consensus on the nature of the concept symbol.

We used content analysis to develop a classification scheme (Table 1) for the content of all citation contexts. We considered only content relating to the material in the 1975 edition of *MMM*. Two pairs of coders independently tested the various versions of the classification scheme; overall inter-coder agreement using the final version was ~90%. Differences were negotiated and, ultimately, some contexts were deemed unclassifiable.

Table 1. Concepts for which *The Mythical Man-Month* is cited

	GENERALIA
CLASS A	Sole characterization of the book as **a classic of software engineering**, etc.
CLASS B	Software development is a "**tar pit**" or "asphalt swamp" [engulfing the programmer]
CLASS C	**Programming is "an art**;" "the programer works with pure 'thought stuff'"
	PROJECT MANAGEMENT ISSUES
CLASS D	Invocations of "**Brooks Law**" [the *mythical man-month*] adding more people to a late project makes it later. Includes problems of coordination and collaboration
CLASS E	**Programmer productivity** issues (e.g. relationship of size & complexity to productivity, quantifying programming effort)
CLASS F	**Managerial concerns** in overall project planning (e.g. overall time and cost estimates)
CLASS G	**General management issues** not specifically related to project planning
	BUILDING THE SYSTEM
CLASS H	**Conceptual integrity** issues, (e.g. the concepts of **Software Architecture** and **Chief Programmer** overseeing a **Surgical Team**)
CLASS I	Quantitative relationship of project scale & complexity to effort expended – **1:3:3:9** (programming in the small to programming in the large)
CLASS J	Prototyping the system – "**Plan to throw one away. You will anyhow**."
CLASS K	Topics in software structure and documentation (not overall aspects of design). Includes the idea that "**the flowchart is a bad/unnecessary thing**; " that **all programs need documentation**, that **modular programming** is good, and that **all software has bugs** and will degrade over time.
CLASS L	The "**second system effect**." [An early version of "feature creep"]
	OTHER
CLASS M	Reference is to Brooks' essay "No Silver Bullet" (which was included in the silver anniversary edition but is also referenced erroneously to earlier editions of *MMM*)
CLASS N	Not classifiable.

Sources of error and bias

The usual caveats apply when the ISI citation indexes are used as data sources. Almost all citations are from journal articles, eliminating the ability to examine influence via citations from conference papers, trade publications, and books—all important to IT professionals in academia and practice. There is a natural bias toward papers published in English which was accentuated by our dropping of a few (less than 10) non-English papers citing *MMM* in the time period covered. The final set of classes are not uniformly homogenous—we tried to strike a balance between (1) establishing separate categories for infrequently encountered ideas and (2) showing shifts of interest, over time and subject area, of topical areas in *MMM*. Overall, however, the classification scheme appears to capture and distinguish between the main ideas that authors recognize in *MMM*.

Results

Due to length restrictions, we focus solely on the aggregate (1975-1999) results in this report. The results of our longitudinal analysis will be reported elsewhere.

Article-level Results

Figure 1 shows the citation history of *MMM* from the first year of publication through the end of data collection in May, 1999. The book, in its various editions, received an average of almost 27 citations/year in this time period with a generally upward, though fluctuating trend. (This trend continued in later years; from Jan 1999 through December 2004 the book received between 33 and 44 citations per year, pointing to its continued popularity.)

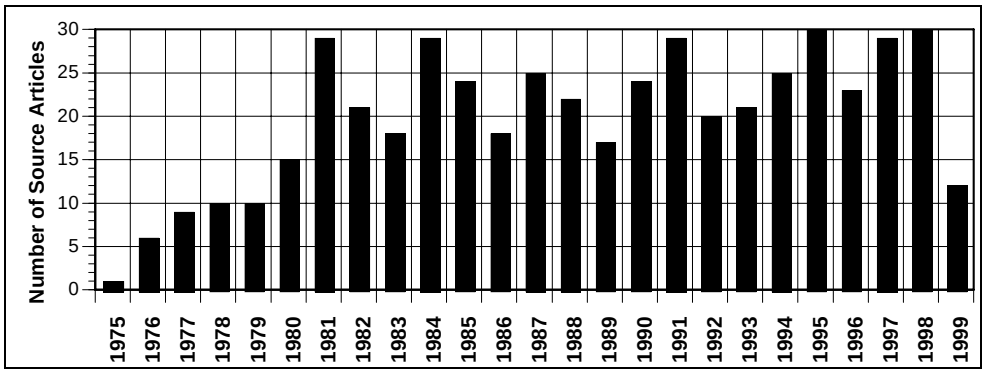

Figure 1: Citation history of *The Mythical Man-Month,* 1975-May 1999

Table 2 gives counts for the 15 aggregate journal subject areas. At the citing article level, Software Engineering and Computer Science (CS) are the two fields in which *MMM* is most heavily cited. They account, along with the recently emerging area of Information Systems (IS), for more than half of the 497 articles analyzed in this study. The appearance of Management (MGT) in the third spot is also not surprising, since many of the issues discussed by Brooks are lessons learned that can apply equally to the management of any large, complex project. The "Other Science" (OS) category includes a variety of journals in physics, chemistry, the life sciences and medicine, along with *Science* and other general science journals. The majority of journal articles in Electrical Engineering (EE) come from *Proceedings of the IEEE.*

Table 2. Subject distribution of citing articles

Journal Subject Area	**Number of Articles**
Software Engineering (SE)	139
Computer Science (less SE and IS)	137
Management and Industrial Engineering	61
Other Sciences (OS)	31
Electrical Engineering (EE)	23
Information Systems (IS)	21
Ergonomics and HCI (Erg/HCI)	12
Other Social Sciences (OSS)	11
Education (Edu)	10
History & Humanities (HH)	10
Library & Information Science (LIS)	10
Psychology (Psyc)	10
Other technology (OT)	10
Law	7
Other Business (OB)	5

Context-level results

Of the 574 citation contexts analyzed, 513 could be assigned to one of the 4 sets of main concept classes (Table 1) relevant to the content of the 1975 edition of *MMM*. The remainder (Classes M and N) were either references to the "no silver bullet" essay (originally published in 1987 and included in the 1995 edition of *MMM*) or were not classifiable. Table 3 shows the distribution of citation contexts over content classes A through L (rows) and (inherited) journal article subject areas (columns) for the aggregate time period 1975-1999. Subject areas are sorted in descending order to emphasize the concentration of citation contexts in a limited number of areas. Row and column percentages report relative concentration in subject areas and content classes (overall percent uniformity). Figure 2 shows the distribution of content classes across the top subject areas.

Table 3. Distribution of content classes across all subject areas

	CONTENT CLASSES													
	D	**A**	**F**	**K**	**H**	**J**	**E**	I	C	B	L	G	**Total**	**% SA**
SE	**36**	**25**	**18**	**16**	7	**19**	6	7	3	2	4	0	**143**	27.9%
CS	**27**	**18**	**19**	**14**	**16**	**11**	**10**	5	3	2	1	3	**129**	25.1%
Mgt	**20**	6	**13**	5	5	7	3	3	1	1	0	0	**64**	12.5%
OS	**10**	**9**	4	4	5	1	4	5	0	1	0	1	**44**	8.6%
IS	**9**	7	2	5	2	2	1	1	0	0	0	1	**30**	5.8%
EE	4	1	3	4	2	1	6	1	0	0	1	0	**23**	4.5%
Erg/HCI	0	0	1	5	3	2	0	0	0	0	0	0	**11**	2.1%
LIS	2	0	2	3	1	0	2	1	0	0	0	0	**11**	2.1%
Edu	1	4	1	0	1	2	1	0	0	0	0	0	**10**	1.9%
OT	1	1	2	0	3	1	1	0	0	0	0	0	**9**	1.8%
Psyc	2	1	2	2	2	0	0	0	0	0	0	0	**9**	1.8%
HH	1	4	1	1	0	0	0	0	1	0	0	0	**8**	1.6%
OSS	1	1	1	2	2	0	0	0	1	0	0	0	**8**	1.6%
Law	2	0	0	0	2	0	1	0	2	0	0	0	**7**	1.4%
OB	4	0	0	0	0	0	2	0	1	0	0	0	**7**	1.4%
Total	**120**	**77**	**69**	**61**	**51**	**46**	**37**	**23**	**12**	**6**	**6**	**5**	**513**	100.0%
% Unif.	.4%	.0%	.5%	.9%	9%	0%	2%	5%	3%	2%	2%	0%	0.0%	

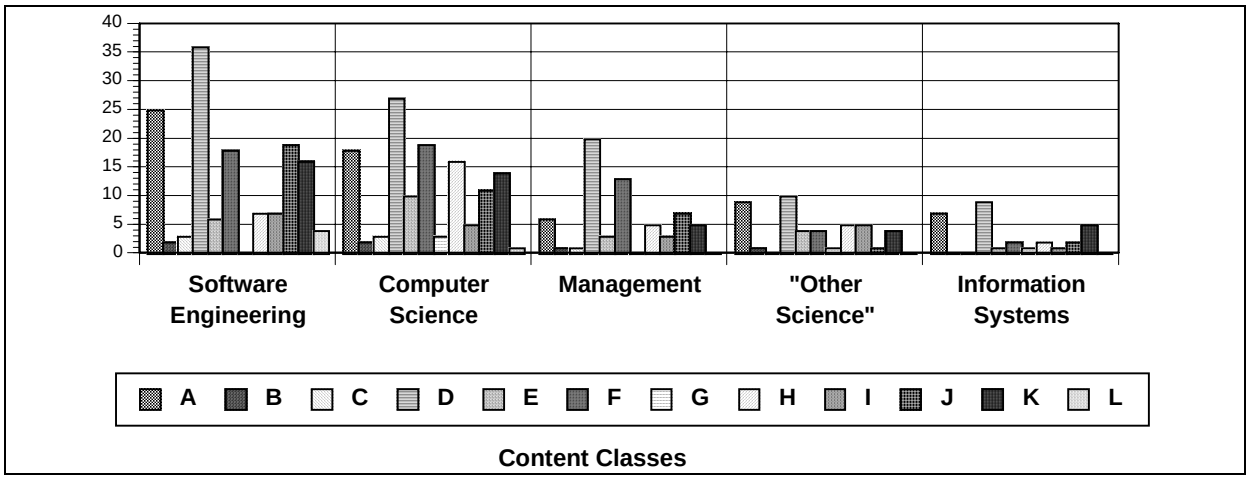

Figure 2: Distribution of content classes in 5 top subject areas

We can see that, for the time period as a whole, slightly over 23% of all classifiable citation contexts point to the eponymous concept. "Brooks' Law" (Class D—adding more people to a late project makes it later). The second most frequently encountered concept class overall (15%) is Class A (*MMM* is a classic of software engineering/an account of managing the development of OS360, etc). This is true for all of the top subject areas except Management, which, as can be seen in Figure 2, focuses secondarily on Class F (overall management concerns in project planning). Computer Science and Software Engineering also include double-digit citations of class J (the usefulness of prototyping—"plan to throw one away, you will anyhow") and Class K (topics in software structure and documentation). While it appears that project management issues (Class D, F) dominate as the most frequently encountered specific concept symbols (as opposed to Class A which is a general characterization) neither has a sufficiently high percent uniformity to identify it as the concept for which *MMM* is primarily cited.

Discussion

Is The Mythical Man-Month a "landmark publication?"
It's clear from the data shown in Figure 1 that Brooks' *The Mythical Man-Month* has "legs." After its initial rise in citation count, it has received between 15 and 30 citations per year over the time period studied (and, as noted earlier, this trend has continued through the end of 2004). *MMM*'s citation history also illustrates the spread and breadth of its influence in terms of the range of citing subject areas. Over the entire period, *MMM* remained highly cited in its "home areas"—Computer Science and Software Engineering—while spreading to other cognate areas such as Management and Information Systems. The raw data also include more than 50 ISI subject categories (condensed to 15 subject areas for analysis) across the sciences, arts, humanities, and areas of application (e.g. medicine and law) in which at least one article citing *MMM* over the full time period.

There are limited comparisons we can make with other studies to evaluate the "landmark" or "citation classic" status of *MMM*. It does show a "type 2" citation history (Aversa, 1985, McCain & Turner, 1989) where the citation rate peaks at year 6-7 or later. (Aversa studied a hyper-cited set of articles in Chemistry, while McCain and Turner (1989) examined core articles dealing with the construction of gene libraries, an area in Molecular Genetics.) Similar citation profiles were reported by Cano and Lind (1991) for 10 ISI "Citation Classics." The curves characterized by Aversa, McCain and Turner, and Cano and Lind were for highly cited journal articles; however, the pattern is also similar to that reported by both Garfield (1985 a, b) and Furner (2003) for Price's *Little Science, Big Science*, arguably a landmark publication in scientometrics, sociology of science and science policy. Additionally, the general distribution is similar in shape and amplitude to those reported for core monographs in philosophy and sociology (though not economics) (Lindholm-Romantschuk & Warner, 1996) and ornithology (Lewison 2004). The subject span of *MMM* is also similar to the "over 80 specialties and disciplines" reported by (Garfield, 1985a, b) for *LSBS*. Additionally, like *LSBS, MMM* received extensive, though sporadic, notice from journals across the academic spectrum. It seems clear from the evidence that *MMM* can be considered a highly influential classic book in software engineering, computer science, information systems and management and a book with substantial insights to offer in fields across the academic spectrum.

If The Mythical man-Month is a concept symbol, what does it symbolize?
MMM appears to represent a number of different concepts to citing authors—not surprising in a book of 195 pages (1975 edition). Our initial rough sorting of citation contexts, based solely on text string similarity, yielded more than 30 groupings and a small pile of apparent singletons. By comparing text content with Brooks' commentary in *MMM*, we established the final set of 15 main classes shown in Table 1. Table 3 shows the overall frequency with which the 15 different concept classes were actually invoked in citation contexts. One concept, Class A, functions as a general reference to the work as a whole— sometimes characterizing the book (*MMM* is a classic of software engineering, a description of the development of IBM's OS360, etc.) and, at other times, simply including it in a bibliography or a broadly described list of works on a topic. The remaining classes focus on specific ideas, discussion points, and lessons learned.

We began this study anticipating that the title concept—*The Mythical Man-Month* (aka Brooks' Law)—would be the major specific concept occurring in citing contexts. To our surprise, less than a quarter of the 513 classifiable contexts referred to Brooks' Law in any form whatsoever when we look at the overall distribution of contexts over classes. The percent uniformity, established by Small (1978) as a measure of the degree to which a work has become a "standard symbol," is 23.4% for *MMM* (Table 3)—roughly a third of the 68% percent uniformity observed by Small for books in chemistry (the journal articles overall had a percent uniformity of 92%).

Additionally, as Figure 2 shows, *MMM* has meant different things to authors in different subject areas. While Classes A and D are prominent in all of the 5 top subject areas, Class F holds second place overall in the Management area and Classes F, J, and K are in a virtual dead heat in Software Engineering. Computer science is the only one of the top 5 with a high count for Class H.

There is little in the literature to which we can directly compare our results concerning the uniformity of citation to *MMM* as one or a set of concept symbols. As noted earlier, Small (1978) reported a distinctly lower percentage uniformity for highly cited books in Chemistry, as opposed to journal articles. While Garfield (1985a, b) did not conduct a formal citation content analysis in his analysis of *LSBS*, he did discuss the diversity of Price's ideas that were cited by a random selection of source articles. Furner (2003) expands on these, again without any quantitative analysis.

It would be a mistake, however, to draw too firm a line between books and journal articles. Journal articles, too, may have "multiple personalities." Cozzens (1982) reports a case of "split identity" – an article in economics that was being cited for two different (though interrelated) concepts by authors associated with two different document co-citation clusters. McCain and Turner (1989) list four different concepts (experimental methods, specific research results, theoretical contributions, existence of research materials) for which a set of molecular genetics papers could be cited. Mizruchi and Fein (1999) report the degree to which empirically-oriented citing authors focused on (and in some cases misrepresented) one or another of three components of a model in a classic paper in organizational theory. Also, Cronin's proposed idea of "tiered citations" suggests that almost any work can be cited at different levels of granularity, from its inclusion in the citation of a complete *oeuvre* to its being cited for a specific method, result, formula, etc. (Cronin, 1994; see also Budd, 1999). In the case of *MMM* the Class A (general) citation contexts correlate to Cronin's "opus" level (citation of a work as a whole) and the remainder (Classes D through L) to the most granular "quantum" level, since they invoke specific ideas, knowledge claims, formulae, etc. in *MMM*.

Conclusions

Brooks' Law is, overall, the most frequently invoked of many different concepts contained in the 1975 edition of *The Mythical Man-Month*. However, we can't say that it is the idea that readers find useful and citeable. It may be the variety as well as the centrality of the many issues and lessons learned in *MMM* that have kept it visible and cited more than 25 years after its initial publication. McCain & Turner and others have noted that one way to maintain a long and healthy citation history is to become a concept symbol for an important methodological contribution. While the "lessons learned" in *MMM* aren't in the same category as the "standard method for protein determination" (Lowry et al., 1951), they can be construed as recommendations for "how to do it right" as well as principles to examine when "things go wrong." In 1999 Verner and colleagues asked "In the 25 years since the Mythical Man-Month what have we learned about project management?"(Verner et al., 1999) The answer then was, when management seems to have paid attention to Brooks' principles, projects were more successful than when they did not. Failures were largely, though not exclusively, attributable to problems in the same areas that Brooks discussed—estimation, planning, communication/organization, risk/change management, and specification issues. We suspect that little has changed since then.

Acknowledgments

We thank Nan Zhou for helping to develop the final version of the classification scheme and for serving as an independent coder. This paper benefited from discussions with our Drexel colleagues in software engineering, but any errors of interpretation are strictly our own.

References

Aversa, E.S. (1985). Citation patterns of highly cited papers and their relationship to literature aging: a study of the working literature. *Scientometrics* 7, 383-389.

Brooks, F.P., Jr. (1975). *The Mythical Man-Month. Essays on Software Engineering.* Reading, MA: Addison-Wesley: Reading, MA.

Brooks, F.P., Jr. (1987). No silver bullet – essences and accidents of software engineering. *Computer* 20(4):10-19.

Brooks, F.P., Jr. (1995). *The Mythical Man-Month. Essays on Software Engineering* Anniversary Edition. Reading, MA: Addison-Wesley.

Budd, J.M. (1999). Citations and knowledge claims: sociology of knowledge as a case in point. *Journal of Information Science* 25, 265-274.

Cano, V. & Lind, N.C. (1991). Citation life cycles of ten citation classics. *Scientometrics* 22, 297-312.

Cronin, B. (1994). Tiered citation and measures of document similarity. *Journal of the American Society for Information Science* 45, 537-538.

Coulter, N., Monarch, I. & Konda, S. (1998). Software engineering as seen through its research literature: a study in co-word analysis. *Journal of the American Society for Information Science* 49, 1206-1223.

Cozzens, S.E. (1982). Split citation identify: A case study from economics. *Journal of the American Society for Information Science* 33, 233-236.

Cozzens, S.E. (1985). Comparing the sciences: Citation context analysis of papers from neuropharmacology and the sociology of science. *Social Studies of Science* 15, 127-153.

Furner, J. (2003). Little book, big book: before and after *Little Science, Big Science:* a review article, Part II. *Journal of Librarianship and Information Science* 35, 189-201.

Garfield, E. (1985a). In tribute to Derek John de Solla Price: a citation analysis of *Little Science, Big* Science. *Scientometrics* 7, 487-503.

_____ (1985b). In tribute to Derek John de Solla Price: a citation analysis of *Little Science, Big* Science. *Current Contents* 24, 232-240.

Hargens, L.L. (2000). Using the literature: reference networks, reference contexts, and the social structure of scholarship. *American Sociological Review* 65, 846-865.

Lewison, G. (2004). James Bond and citations to his books. *Scientometrics* 59, 311-320.

Lindholm-Romantschuk, Y. & Warner, J. (1996). The role of monographs in scholarly communication: An empirical study of philosophy, sociology, and economics. *Scientometrics* 52, 389-404.

Lowry, O.H., Rosbrough, N.J., Farr, A.L. & Randall, R.J. (1951). Protein measurement with the folin phenol reagent. *Journal of Biological Chemistry. 193,* 265-275.

Marion,L.S. & McCain, K.W. (2001). Contrasting views of software engineering journals: Author cocitation choices and indexer vocabulary assignments. *Journal of the American Society for Information Science* 52, 297-308.

McCain, K.W. & Turner, K. (1989). Citation content analysis and aging patterns of journal articles in molecular genetics. *Scientometrics* 17, 127-163.

McCain, K.W., Verner, J.M., Hislop, G.W., Evanco, W.M, & Cole, V. (2003). Combining bibliometric and knowledge elicitation techniques to map a knowledge domain. *Mapping Knowledge Domains.* Arthur M. Sackler Colloquia of the National Academy of Sciences. May 9-11, 2003. Irvine, CA.

Mizruchi,M.S. & Fein, L.C. (1999). The social construction of organizational knowledge: A study of the uses of coercive, mimetic, and normative isomorphism. *Administrative Science Quarterly* 44, 653-683.

Small, H.G. (1978). Cited documents as concept symbols. *Social Studies of Science* 8, 327-340

Small, H. (1982). Citation context analysis. In B. Dervin & M.J. Voight (Eds), *Progress in Communication Sciences* 3 (pp.287-310). Norwood, NJ: Ablex.

Small, H. & Greenlee, E. (1980). Citation context analysis of a co-citation cluster: Recombinant DNA. *Scientometrics* 2, 277-301.

Verner, J.M., Overmyer, S.P. & McCain, K.W. (1999). In the 25 years since *The Mythical Man-Month* what have we learned about project management? *Information and Software Technology* 41, 1021-1026.

Verner, J.M., Evanco, W.M., McCain, KW, Hislop, G.W. & Cole, V. 2001. The determinants of visibility of software engineering researchers. *Journal of Systems and Software* 59, 99-106.