

# Citation Analysis of Ph.D. Dissertation References as a Tool for Collection Management in an Academic Chemistry Library

Núria Vallmitjana and L. G. Sabaté

A bibliometric study was carried out on the citations within the chemistry field Ph.D. dissertations to ascertain what types of documents are the most frequently used in the research process, the most frequently consulted journals and obsolescence rate of the journals. The analysis covered 46 doctoral theses presented at the Institut Químic de Sarriá (IQS) from 1995 to 2003. The results obtained from the 4,203 citations revealed that the most frequently used documents were scientific papers, which accounted for 79 percent of the total; 33 journals met 50 percent of the informational needs; and the age of 50 percent of the citations was no older than 9 years. Finally, the results can be used as a tool for the collection management of the library.

## About the IQS

The Institut Químic de Sarriá (IQS) of the Universitat Ramon Llull is an academic centre founded by the “Compañía de Jesús” in 1905 and located in Barcelona, Spain. In the last hundred years, teaching and research in Chemistry and Chemical Engineering has been its main interest. Its Chemical Engineering course has been ABET-accredited since September 2004. Its Ph.D. courses in Chemistry and Chemical Engineering received the Quality Mention from Spanish Government in 2003.

## Introduction

It seems paradoxical that, in the information age, it could be so difficult to have all the journals published in a specific scientific field available. Some of the main reasons are: the great number of scientific journals published, their high subscription cost, with yearly price increases that are significantly higher than the standard of living increase, and the budgetary restraints of academic libraries. These reasons force the libraries to establish priorities in their collection acquisitions and maintenance policy in accordance

---

*Núria Vallmitjana is associated with the Library at the Institut Químic de Sarriá (Universitat Ramon Llull); e-mail: nuria.vallmitjana@iqs.edu. L. G. Sabaté works in the Department of Applied Statistics, Facultat d'Economia IQS (Universitat Ramon Llull); e-mail: lgsab@iqs.edu. The authors would like to thank the students of the 2001–2005 Business Administration and Management at the IQS for analyzing most of the theses as part of a joint practical work for both “Information Management” and “Statistics” courses. Our thanks also go to Martí G. Gatell, who developed the database that was used in this educational activity, and X. Tomás for his useful comments.*

with the users' needs. The users' needs can be identified by the analysis of the scientific literature usage of students, teachers, and researchers.<sup>1</sup>

Journal usage can be estimated by the number of citations contained in the documents published by researchers as the articles in scientific journals, conference papers or postgraduate dissertations. These documents contain a great number of bibliographical citations because "students tend to be exhaustive and chronologically complete in the review of the literature."<sup>2</sup>

The analysis of the Ph.D. dissertation citations constitutes a good method for the evaluation of journal usage, because it is the best way of identifying the sources of information looked up by the researchers and, therefore, justifies the investment devoted to the subscriptions.<sup>3</sup>

This work studies the bibliographical references of the Ph.D. dissertations in Chemistry, as a source of information to manage the scientific journal collection in the academic library. The analysis answers the following questions:

- What is the proportion of journals cited?
- What are the most cited journals?
- Is there any relationship between the top journals' rank and their impact factor?
- How old are the cited articles?
- How much is the cost per citation?

The results can be used as management criteria for the literature collection of a library. The results let us decide about renewals, subscriptions, and cancellations of the journals. It will also help us to manage the situation of the journals in the available space by their obsolescence rate.

The revision of the literature we present shows that there are few works published in this field and all have been carried out in libraries of the United States, and this is the first and only work that refers to a European centre.

The bibliometric analysis carried out in this study uses a methodology similar

to the one used by Chrzastowski<sup>4</sup> in 1991 and by Gooden<sup>5</sup> ten years later. Both works are centered on Chemistry and come from data obtained in a local-use study. Nevertheless, the first study uses four parameters (the reshelving of all journals picked up throughout the library each day, use through in-house circulation, journal lending and borrowing, and through interlibrary loan). On the other hand, the second study uses the citations of the Ph.D. dissertations as a frequency indicator of journal use.

### Literature Review

The analysis of the citations cited in the Ph.D. dissertations has been the tool to evaluate the researchers' information needs and the role of libraries to satisfy those needs. Kushkowsky<sup>6</sup> organizes in a table some of the most important works published in the last years in different areas of knowledge. In all of the works there is an increase of citations as time goes by, although the number of citations changes in each discipline, and it is recommended to study monographically this trend in each of these areas.

The bibliometric study of Buchanan and Herubel<sup>7</sup> in Political Science concludes that scientific journals represent the biggest proportion of materials cited in the dissertations and that the materials cited in these documents and the analysis of them allow the library a better management of the collection.

Sylvia and Lesher<sup>8</sup> apply the method of analysis of citations from theses and dissertations to the field of Psychology. Their study evaluates two parameters that are complementary: the number of journals that are most frequently cited and their *cost-per-use*.

Zipp<sup>9</sup> not only analyzes the citations of the dissertations to evaluate the collection of the library specialized in Geology, but also studies the citations of the articles published by the researchers and compares both results.

Buttlar<sup>10</sup> analyses the citations of Ph.D. dissertations on Library Science and

Information Science for the evaluation of the nature of material cited most, the authors, the countries of origin of publications cited, the journals cited, their range of topics and how current is the literature cited.

Smith<sup>11</sup> studies the results of the citations analysis of a sample of theses from 2001 about several different topics and compares them with the analysis realized ten years before. He evaluates the usefulness of the library collection and investigates its evolution through time, in accordance with the introduction of electronic information sources and with the massive increase in the cost of subscriptions to scientific journals. Arts and Humanities tend to depend more on monographs; meanwhile, Science and Technology use mostly scientific journals. The age of the cited materials also establishes several differences: Science and Technology seem not to be interested in old material like Arts and Humanities.

Haycock<sup>12</sup> analyzes the citations in dissertations on Education Sciences, finished between 2000 and 2002 to establish which are the journals that were most frequently used as well as the relationship between monographs and articles of scientific journals.

Gooden<sup>13</sup> analyzes the Ph.D. dissertations of the Ohio State University Chemistry Department between 1996 and 2000. The conclusion reached is that only 12 journals are necessary to cover 50 percent of the references and demonstrates that most of the citations correspond to articles published in scientific journals.

Another work in the Chemistry field is one of Chrzastowski<sup>14</sup> that also studies the use of journals in the library of the University of Illinois at Urbana-Champaign (UIUC). She calculates the cost-effectiveness of the collection to establish which subscriptions should be cancelled considering that the relationship price/use is too high. In comparison with the aforementioned works, the use of the collection is measured with four indicators (the reshelving count of all journals

picked up throughout the library each day, the number of issues used through in-house circulation, the lending and borrowing journal count and also through interlibrary loan). In this work, the journal usage is not measured by means of the citations contained in the Ph.D. dissertations. Chrzastowski and Olesko<sup>15</sup> continue with the former study and analyze the trends of use and cost of the journals in the UIUC Chemistry Library.

### Methodology

The starting material for this work is the front page of each thesis and the literature cited.

The information registered for each thesis is:

- identification data, including the year;
- the number of documents cited: monographs, theses and academic dissertations, articles, and other documents;
- for each cited article there is an identification code (theses-article), the journal title and the publication year.

Data has been processed in the following way:

- citation analysis according to document type;
- bibliometric analysis of the cited journals: frequency distribution and impact factor;
- description of the citation age distribution;
- cost per citation study.

### Results

From 1995 to 2003, 68 Ph.D. theses were presented in the Institut Químic de Sarrià (IQS), in Chemistry or Chemical Engineering fields. Twenty-two theses were put aside because they presented the bibliography in chapters or because they were not available. The results obtained from the 46 theses analyzed are as follows:

#### *What is the Proportion of Journals Cited?*

According to the document type, 79 percent are articles in scientific journals, 12

percent are monographs, and 2 percent are theses or academic dissertations.

Citations numbering 4,203 are articles published in 593 journals. The relationship between the number of cited articles and the number of cited journals is 7.1. The average number of scientific articles cited per theses is 91.

**What are the Most Cited Journals?**

In accordance with Bradford’s<sup>16</sup> Law, most of the articles on a certain subject are published in a reduced number of journals. The analysis of the studied references reveals that a core formed by 33 journals satisfies 50 percent of the informational needs and that 150 journals are needed to reach 80 percent of the informational necessities as shown in Figure 1.

Table 1 lists the 33 journal titles needed to satisfy 50 percent of the journal citations in this study.

From the 33 titles that constitute the core journals, 14 belong to Organic Chemistry and Biochemistry, 9 to General Chemistry; meanwhile, the rest are distributed in several areas such as Environmental Science, Electrochemistry, and Corrosion among others.

Taking into account the publisher of this core journal, 58 percent of titles come from three publishing firms:

- 9 belong to Elsevier (27%);
- 6 belong to the American Chemical Society (18%);
- 4 belong to Wiley-VCH (12%).

For the rest of the journals, 12 have been published by different scientific societies and only 2 belong to commercial publishing firms (Nature Publishing and Thieme).

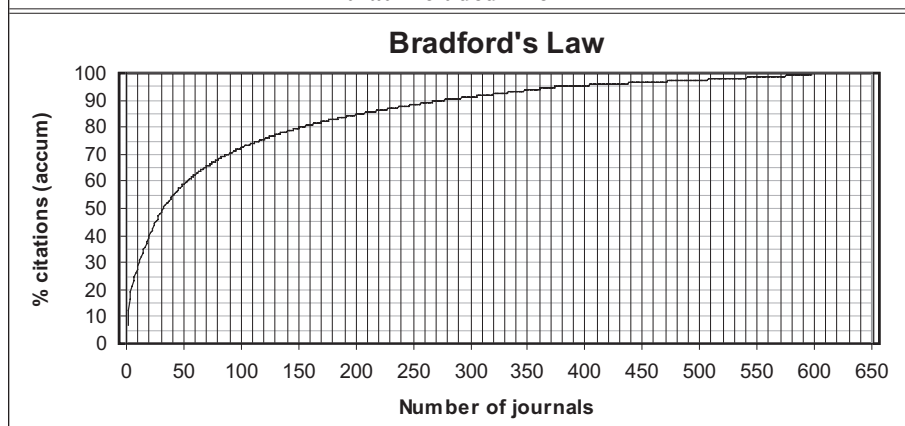
**Is There Any Relationship Between the Top Journals’ Rank and Their Impact Factor?**

The impact factor for 2003 of the top journals obtained from *Journal Citation Reports* varies from 0.157 to 30.979. Figure 2 shows that there exists a moderate association between the position of a core journal and its impact factor rank measured with the correlation coefficient of Spearman ( $r_s=0.5194$ ;  $p\text{-value}=0.0033$ ). This possible association must be studied more in-depth.

**How Old are the Cited Articles?**

The reference age is defined as the time passed between the publication year of an article and the thesis reading year where this paper is cited. Figure 3 shows the frequency distribution of citation ages. The mean is 14 years with a median of 9 years, while 90 percent of citations are 31 years old or fewer; however, the range is 145 years.

**FIGURE 1**  
Citations Accumulated According to the Number of Journals that Included Them



**TABLE 1**  
**Top 33 Core Journals Cited**

<b>Rank Core Journal</b>	<b>Citations (%)</b>	<b>Accum. Citations (%)</b>	<b>Journal Title</b>	<b>Impact Factor 2003</b>	<b>Rank Impact Factor 2003</b>	<b>Cost per Citation (€)</b>
1	6.87	6.87	Journal of the American Chemical Society	6.516	4	0.20
2	5.28	12.15	Journal of Organic Chemistry	3.297	9	0.18
3	4.60	16.75	Tetrahedron Letters	2.326	15	1.09
4	2.52	19.27	Tetrahedron	2.641	12	2.27
5	2.03	21.30	Angewandte Chemie. International Edition in English	8.427	3	0.85
6	1.78	23.08	Chemosphere	1.904	22	1.08
7	1.47	24.55	Biochemistry	3.922	7	1.05
8	1.43	25.98	Journal of Medicinal Chemistry	4.820	5	0.55
9	1.36	27.34	Journal of the Electrochemical Society	2.361	13	0.26
10	1.31	28.65	Afinidad	0.157	31	0.01
11	1.26	29.91	Tetrahedron: Asymmetry	2.178	16	0.90
12	1.24	31.15	Environmental Science & Technology	3.592	8	0.51
13	1.21	32.36	Helvetica Chimica Acta	1.861	23	0.69
14	1.21	33.57	Journal of Biological Chemistry	2.361	14	0.71
15	1.14	34.71	Carbohydrate Research	1.533	26	2.71
16	1.12	35.83	Nature	30.979	1	0.60
17	1.07	36.90	Chemical Reviews	21.036	2	0.38
18	1.07	37.97	Journal of Chromatography	2.922	11	5.22
19	0.96	38.93	Journal Chemical Society. Perkin Transactions I	1.948	21	—
20	0.93	39.86	Syntheses	2.074	18	0.62
21	0.91	40.77	Biotechnology and Bioengineering	2.173	17	2.31
22	0.91	41.68	Electrochimica Acta	1.996	20	2.10
23	0.91	42.59	Heterocycles	1.082	29	2.02

**TABLE 1**  
**Top 33 Core Journals Cited**

Rank Core Journal	Citations (%)	Accum. Citations (%)	Journal Title	Impact Factor 2003	Rank Impact Factor 2003	Cost per Citation (€)
24	0.91	43.50	Journal Chemical Society	—	32	—
25	0.91	44.41	Journal Chemical Society. Chemical Communications	4.031	6	0.80
26	0.86	45.27	Chemistry Letters	1.579	24	0.20
27	0.84	46.11	Chemical Engineering Science	1.562	25	2.66
28	0.84	46.95	Corrosion	0.774	30	0.17
29	0.77	47.72	Journal of Organometallic Chemistry	2.042	19	6.21
30	0.75	48.47	Bulletin of the Chemical Society of Japan	1.237	27	0.45
31	0.75	49.22	Chemische Berichte	—	33	—
32	0.75	49.97	Journal of Chemical Physics	2.950	10	3.47
33	0.65	50.62	Canadian Journal of Chemistry	1.157	28	0.68

#### *How Much is the Cost Per Citation?*

The last aspect considered is the economic cost of the core journals. If we take the subscription cost for 2004, the library must have spent about 100,000 € to maintain the subscriptions of the core journals.

At the same time, considering the 46 theses studied, the subscription price of the core journals and the number of citations for each journal, the average cost is about 1 € / citation; 57 percent of the journals have a cost lower than 1.00 €/citation; meanwhile, 14 percent have a cost higher than 3.00 €/citation (see table 1).

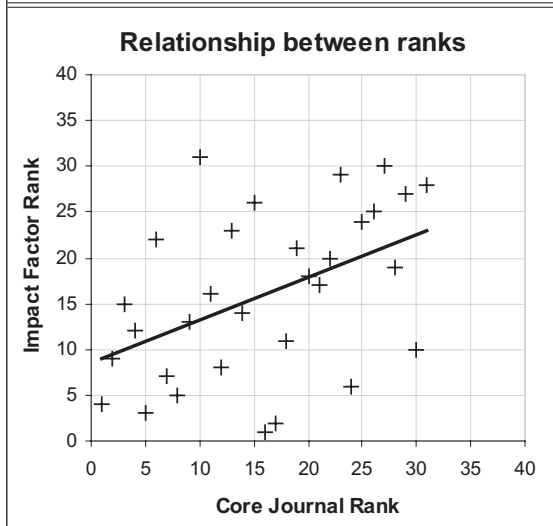
#### **Discussion**

As highlighted in table 2, the results obtained in this work (Vallmitjana column) are comparable to the ones of other works cited before.

All the works included in the table were published by different authors between the years of 1994 and 2004. The methodology used is based on the bibliometric analysis of the references contained within doctoral theses from different universities in the United States. Most of the studies are centered in the thesis of an specific subject except for the multidisciplinary studies of Kushkowski and Smith.<sup>17</sup> The chronological timeframe is also variable; some studies cover Ph.D. dissertations from a specific year, whereas other ones cover theses from several years. The proportion of references analyzed also varies; in some cases, all the references are analyzed, whereas in others only a representative sample is studied.

In the mentioned works, one can conclude that the number of references contained within doctoral theses increases with time; the authors attribute this to the fact that information is more and more easily accessible. Also, they conclude that the average number of references by thesis varies between subjects. Social Sciences and Arts and Humanities use fewer references than those of Natural Sciences (Kushkowski, Smith<sup>18</sup>). These facts are reflected in the comparison of results in table 2. The work of Gooden,<sup>19</sup> in the field of Chemistry, states that 86 percent of references were scientific papers published in journals. Kushkowski<sup>20</sup> indicates that the percentage of journals used in Biology and Physics also surpassed 80 percent. For Smith,<sup>21</sup> in the field of Sciences, the percentage was also 80 percent; Arts and Humanities are those that use 20 percent fewer journals, whereas the percentage in Social Sciences is situated between 40 and 60 percent.

**FIGURE 2**  
**Relationship Between the Core Journals' Rank and their Impact Factor Rank**

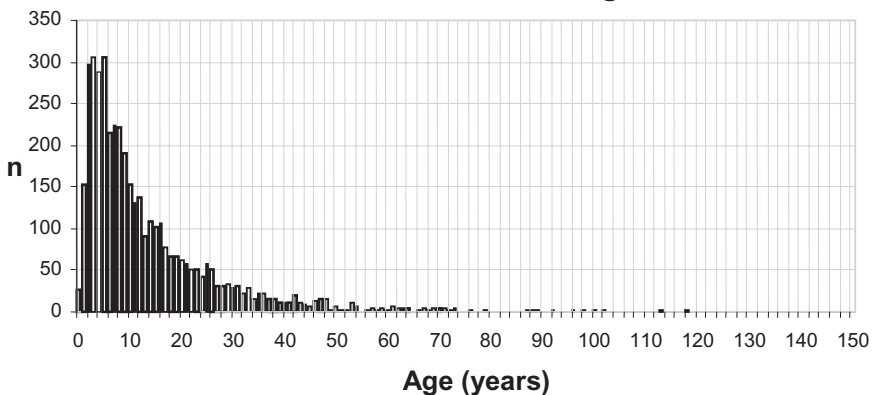


Since only the work of Gooden<sup>22</sup> is in the area of Chemistry, we compared in table 3 some relevant results of her work and ours.

It is important to mention that the number of journals needed to cover 50 percent of journal citations is very different between the Gooden study and our study. Gooden establishes a group

**FIGURE 3**

**Distribution of citation ages**



**TABLE 2**  
**Results Comparison of Several Works**

	Buchanan	Buttlar	Gooden	Haycock	Kushkowski	Smith	Vallmitjana (This paper)
Country	USA	USA	USA	USA	USA	USA	Spain
Area	Political Sci.	Lib. & Inf. Sci.	Chemistry	Educ. Sci.	Multidisciplin.	Multidisciplin.	Chemistry
Period	1979–1989	1994–1997	1996–2000	2000–2002	1973–1992	2001	1995–2003
No. of theses	32	61	30	43	629	30	46
No. of citations	3,673	3,683	3,704	4,542	9,100	1,595	5,320
% articles	37	46	86	44	64	48	79
No. of journals	327	815	441	558	N/A	N/A	593

of 12 titles; meanwhile, in our study the core journals is composed of 33 titles. Gooden determined that 20 titles included 61 percent of the journal references, whereas in our study the 20 most used titles cover only 40 percent of the journal references.

Comparing the titles of the core journals of both studies, it has been observed that:

- 8 titles appear in both lists: *Journal of the American Chemical Society*, *Journal of Chemical Physics*, *Tetrahedron Letters*, *Biochemistry*, *Journal of Organic Chemistry*, *Journal of Biological Chemistry*, *Nature*, and *Journal of the Chemical Society*.

- *Journal of the American Chemical Society* is the most cited in both cases and has been cited 11.45 percent in the Gooden's study and 6.87 percent in the present study.

With regard to the age of the cited documents, the mean age was 14 years with a median value of 9; meanwhile, in the multidisciplinary study of Kushkowski,<sup>23</sup> the mean was 12 years and the median value 8.

Citation costs obtained in this study are not comparable with the ones of Chrzastowski and Olesko,<sup>24</sup> because of the differences in the methodologies employed.

On average, 91 papers are cited in each thesis, the cost per thesis amounts to approximately 90 €. In contrast, if all articles had been acquired through one of the main document suppliers, this would represent a cost of nearly 1,200 € per thesis.

## Conclusions

A bibliometric study was designed in accordance with scientific literature to study the information needs of Ph.D. students.

A total of 4,203 citations in 593 journals were analysed. The analysis accounted for approximately 79 percent of the total citations that were made in 46 of the 68 Ph.D. dissertations accepted by the IQS (Institut Químic de Sarrià, Universitat Ramon Llull) from 1995 to 2003.



**TABLE 3**  
**Comparison of Results**

		Results	Gooden	Vallmitjana (this paper)
Type of document	Journals		86%	79%
	Monographs		8%	12%
	Theses		6%	2%
Total journal citations (A)			3,178	4,203
Number of different journals (B)			441	593
Ratio A/B			7.2	7.1
Number of Theses (C)			30	46
Mean = A/C			106	91
Number of journals needed to cover 50% of journal citations			12	33
Percentage of the journal citations covered by the first 20 titles (Note: we use 20 titles to compare with Gooden's paper)			61%	40%

From the analysis, we conclude that:

- Scientific journals are the most frequently used document type in IQS chemistry Ph.D. dissertations. On average, 91 articles were cited in each thesis.
- 50 percent of the citations come from 33 core journals, which is less than 6 percent of the total number of journals cited.
- A large proportion of the most cited journals come from a small set of publishing companies. Nearly 60 percent of the core journals belong to 3 publishers.
- There is some evidence of a relationship between the impact factor rank and the rank of the journal in terms of citation frequency. This possible relation must be studied more in-depth.
- The age of 50 percent of the citations is no higher than 9 years, although the mean age is 14 years.

No library has unlimited resources allowing it to subscribe to all the journals that their users request. Therefore, the library must draw up an acquisitions policy that concurs with the real possibilities and priorities in research areas.

We propose three criteria that a library may use in deciding which subscriptions to cancel and which to keep. The criteria concludes the following:

- In agreement with the titles ranking, core journals must have priority because they are the most cited.
- The citation cost indicator will let us establish a second level of priorities. Subscriptions with a lower cost-per-citation may have preference to maximize the resources assigned to journal acquisition.
- A reduced number of companies publish a high number of journals. Gathering the titles per publisher is a means of assessing the financial viability of subscribing to individual titles or packages, in either printed or electronic format.

Last, having already developed an analysis model, we can continue monitoring the situation merely by updating the data contained in forthcoming Ph.D. theses. Discovering trends in how specific academic journals are used can help the library to get the most out of its budget.

### Notes

1. J. Thomson and R. Carr, *An Introduction to University Library Administration* (London: Bingley, 1987).

## Citation Analysis of Ph.D. Dissertation References as a Tool 81

2. C.A. Barry, "Information Skills for an Electronic World: Training Doctoral Research Students," *Journal of Information Science* 23, no. 3 (1997): 225-38.
3. T.E. Chrzastowski, "Journal Collection Cost-Effectiveness in an Academic Library: Results of a Cost/Use Survey at the University of Illinois at Urbana-Champaign," *Collection Management* 14, no. 1/2 (1991): 85-98.
4. Ibid.
5. A.M. Gooden, "Citation Analysis of Chemistry Doctoral Dissertations: An Ohio State University Case Study," *Issues in Science and Technology Librarianship* no. 32 (Fall) (2001).
6. J.D. Kushkowski, K.A. Parsons, and W.H. Wiese, "Master's and Doctoral Theses Citations: Analysis and Trends of a Longitudinal Study," *Libraries and the Academy* 3, no. 3 (2003): 459-79.
7. A.L. Buchanan and J.P.V.M. Herubel, "Profiling Phd Dissertation Bibliographies: Serials and Collection Development in Political-Science," *Behavioral & Social Sciences Librarian* 13, no. 1 (1994): 1-10.
8. M. Sylvia and M. Lesher, "What Journals Do Psychology Graduate-Students Need: A Citation Analysis of Theses References," *College & Research Libraries* 56, no. 4 (1995): 313-18.
9. L.S. Zipp, "Theses and Dissertation Citations As Indicators of Faculty Research Use of University Library Journal Collections," *Library Resources & Technical Services* 40, no. 4 (1996): 335-42.
10. L. Buttlar, "Information Sources in Library and Information Science Doctoral Research," *Library & Information Science Research* 21, no. 2 (1999): 227-45.
11. E.T. Smith, "Assessing Collection Usefulness: An Investigation of Library Ownership of the Resources Graduate Students Use," *College & Research Libraries* 64, no. 5 (2003): 344-55.
12. L.A. Haycock, "Citation Analysis of Education Dissertations for Collection Development," *Library Resources & Technical Services* 48, no. 2 (2004): 102-06.
13. Gooden, "Citation Analysis."
14. T.E. Chrzastowski and B.M. Olesko, "Chemistry Journal Use and Cost: Results of a Longitudinal Study," *Library Resources & Technical Services* 41, no. 2 (1997): 101-11.
15. Ibid.
16. B.C. Brookes, "Bradford's Law and the Bibliography of Science," *Nature* 22 (2006): 953-56.
17. Kushkowski et al., "Master's and Doctoral Theses Citations"; Smith, "Assessing Collection Usefulness."
18. Kushkowski et al., "Master's and Doctoral Theses Citations"; Smith, "Assessing Collection Usefulness."
19. Gooden, "Citation Analysis."
20. Kushkowski et al., "Master's and Doctoral Theses Citations."
21. Smith, "Assessing Collection Usefulness."
22. Gooden, "Citation Analysis."
23. Kushkowski et al., "Master's and Doctoral Theses Citations."
24. Chrzastowski and Olesko, "Chemistry Journal Use and Cost."

# TAP INTO A NEW WORLD OF ONLINE REVIEWS



ALL IN THE PALM OF YOUR HAND.

- Access to over 120,000 *Choice* Reviews
- Advanced Search Capabilities
- Enhanced Customization Possibilities
- Download, Print, or E-mail Search Results
- Exclusive Online Content
- Comprehensive Content
- User-Friendly Interface

To learn more about *Choice Reviews Online*, visit  
<http://www.cro2info.org/>

To sign up for free 60-day trial, visit <http://www.cro2.org/>

# CHOICE

CURRENT REVIEWS FOR ACADEMIC LIBRARIES